# THESIS / THÈSE

**DOCTOR OF SCIENCES**

## Development of an innovative quantitative method for allergen detection in food using isotopic dilution mass spectrometry

Gavage, Maxime

*Award date:*
2020

*Awarding institution:*
University of Namur

University de Namur – Faculty of Sciences – Department of Biology
Namur Research Institute for Life Sciences (NARILIS)
Laboratory of Cellular Biochemistry and Biology (URBC)
Rue de Bruxelles, 61 – 5000 Namur, Belgium

CER Groupe – Analytical Laboratory & Immunobiology Laboratory
Rue du Point du Jour, 8 – 6900 Marloie, Belgium & Rue de la Science, 8 – 6900 Aye, Belgium

# Development of an innovative quantitative method for allergen detection in food using isotopic dilution mass spectrometry

Original dissertation presented by Maxime GAVAGE in order to obtain the degree of Doctor in Sciences

September 2020

**Composition of the jury:**

**Prof. Thierry ARNOULD (Promotor)**
UNamur, Narilis – Laboratory of Cellular Biochemistry and Biology – Namur, Belgium

**Prof. Patsy RENARD (Co-promotor)**
UNamur, Narilis – Laboratory of Cellular Biochemistry and Biology – Namur, Belgium

**Prof. Martine RAES (President)**
UNamur, Narilis – Laboratory of Cellular Biochemistry and Biology – Namur, Belgium

**Prof. Fabrice BUREAU**
ULiège, GIGA – Cellular and Molecular Immunology – Liège, Belgium

**Prof. Bruno DE MEULENAER**
UGent – Department of Food technology, Safety and Health – Gent, Belgium

**DVM. Philippe DELAHAUT**
CER Groupe – Marloie, Belgium

**Prof. Gauthier EPPE**
ULiège – Laboratory of Mass Spectrometry – Liège, Belgium

**PhD. Nathalie GILLARD (Scientific Partner)**
CER Groupe – Analytical Laboratory – Marloie, Belgium

**Prof. Georges LOGNAY**
ULiège, Gembloux Agro-Biotech – Laboratory of Analytical Chemistry – Gembloux, Belgium

**Summary**

Food allergen analysis is an essential tool in the development of a risk-based approach for allergen management. Robust, specific and sensitive detection methods are still needed to protect allergic patients and guarantee correct food labelling.

During the last decade, mass spectrometry became the method of choice for allergen analysis. This approach is predominantly performed by specific analysis of peptides obtained by an enzymatic digestion of the proteins of the sample, including the proteins of the allergenic ingredients.

Here, a strategy to develop a food allergen quantitative analysis method was proposed. This strategy was applied to elaborate an UHPLC-MS/MS method for the simultaneous detection and quantification of four allergenic ingredients (egg, milk, peanut and hazelnut) in processed food products.

Quantification strategy was based on isotope dilution considering an original approach using a stable isotope-labelled concatemer as internal standard. Concatemer is an artificial protein, recombinantly produced, and assembling peptide biomarkers from different proteins. Submitted to the enzymatic digestion step, concatemer could represent a relevant internal standard in food allergen analysis. It overcomes limitations of the traditionally used synthetic peptides, while combining advantages of labelled proteins, theoretically ideal but financially unaffordable, for a routine method.

Potential peptide biomarkers were first identified with an experimental approach based on high resolution mass spectrometry. The four considered allergenic ingredients were submitted to different representative food processing techniques and analysed by high resolution mass spectrometry. Among the hundreds of identified peptides, potential peptide biomarkers were selected using a set of selection criteria to ensure the specificity, sensitivity and robustness of the future quantitative method. Ideal peptide biomarkers had to be specific to considered allergenic ingredients, belong to abundant proteins, and be robust to food processing but not prone to missed cleavages by the protease used or to amino acid modifications.

The list of the 55 identified potential peptide biomarkers was reduced during the development of the UHPLC-MS/MS method. Selection was based on sensitivity and selectivity criteria and finally, 19 peptide biomarkers were kept in the method.

These 19 peptide biomarkers were combined in a $^{15}N$ stable isotope-labelled concatemer. Design and labelling strategy of the concatemer were optimized to reach high production yield with a sufficient isotopic enrichment. These conditions were necessary to set up a cost-effective method when compared to the synthetic peptide internal standards and avoid any introduction of false positive results, respectively.

The developed UHPLC-MS/MS was validated using different food matrices incurred or spiked with the four allergenic ingredients. Performance parameters including selectivity, LOD, LOQ, linearity, trueness and precision were evaluated.

The use of the designed concatemer as internal standard was finally compared to synthetic peptides and labelled protein approaches.

In conclusion, the combination of mass spectrometry and stable isotope dilution using concatemer allowed the development of a method for the simultaneous quantification of egg, milk, peanut and hazelnut, four major allergenic ingredients, in processed food products.

**Acknowledgement**

Durant ces quatre années de thèse, j'ai eu la chance de travailler dans différents laboratoires et d'être épaulé par de nombreuses équipes. Je souhaite remercier toutes ces personnes qui, de près ou de loin, m'ont apporté leur aide dans la concrétisation de ce projet.

Cette belle histoire a commencé lorsqu'à la suite de mon mémoire au sein du laboratoire analytique du CER Groupe, les Docteurs Philippe Delahaut et Nathalie Gillard m'ont proposé de prendre part au projet Allersens. Je tiens à les remercier pour la confiance qu'ils m'ont accordée.

Durant ces quatre années, j'ai eu la chance et l'honneur d'être encadré par le Professeur Thierry Arnould, promoteur de ma thèse, le Professeur Patsy Renard, co-promotrice de ma thèse, ainsi que par le Docteur Nathalie Gillard. Je vous remercie pour votre encadrement permanant, votre exigence, vos conseils et vos encouragements qui m'ont permis de progresser, de me surpasser et de mener à bien cette thèse. Au-delà des aspects scientifiques, je tiens à vous remercier du fond du cœur pour votre encadrement qui m'a apporté un réel développement personnel.

C'est au sein de l'équipe DYSO de l'Unité de recherche en biologie cellulaire animale de l'Université de Namur que ma thèse a débuté. Je souhaite remercier toute l'équipe pour le chaleureux accueil qui m'a été offert. Merci à Catherine, Maud et Antoine pour leur soutien dans le laboratoire. Je souhaite tout particulièrement remercier Marc Dieu pour son encadrement en spectrométrie de masse, la confiance qu'il m'a accordée ainsi que pour son humour et sa bonne humeur perpétuelle. Marc, ce fut un réel plaisir de travailler à tes côtés.

J'ai ensuite poursuivi mon travail au sein du département d'immunobiologie du CER Groupe, encadré par le Docteur Patrice Filée. Merci pour tes précieux conseils et tes encouragements. Je tiens également à remercier Christian pour son aide au laboratoire ainsi que toute l'équipe pour leur bonne humeur. Je garderai un excellent souvenir de ces moments passés avec vous.

Mon parcours s'est achevé là où tout a commencé, dans le laboratoire analytique du CER Groupe. Je tiens à remercier toute l'équipe de m'avoir si bien accueilli et de toujours avoir pu compter sur votre soutien. Je tiens particulièrement à remercier Mélanie qui m'a précédé dans l'analyse des allergènes. Merci pour tes conseils, tes encouragements et ta gentillesse. J'ai été très heureux de travailler à tes côtés. Merci à Jean pour son appui HRMS qui a permis de donner une dimension supérieure à ce travail ainsi qu'à sa merveilleuse compagnie dans notre bureau.

Je tiens à remercier l'ensemble des membres de mon jury d'avoir accepté d'évaluer ce travail. Merci pour le temps que vous y avez consacré et merci pour vos conseils lors de la défense privée. Je tiens tout particulièrement à remercier le Professeur Martine Raes, présidente de mon jury, pour sa relecture minutieuse et le temps que nous avons passé ensemble qui ont permis de magnifier ce manuscrit.

Cette thèse faisait partie du projet Allersens, réalisé en collaboration avec l'Institut flamand de recherche pour l'agriculture, la pêche et l'alimentation (ILVO). Je tiens à remercier tous les membres de l'équipe qui ont participé à ce projet et en particulier Kaatje, également doctorante dans ce projet. Il m'aurait été impossible d'obtenir les résultats présentés dans cette thèse sans cette collaboration. Je tiens également à remercier le Professeur Kris Gevaert pour ses précieux conseils lors des différentes publications.

Merci à vous tous pour votre aide durant ces quatre années.

À côté de cette fabuleuse collaboration scientifique, je tiens également à remercier toutes les personnes ayant joué un rôle indirect et qui ont fait de cette thèse, quatre merveilleuses années. Je pense notamment à mes compagnons de la formation U2ES, à mes camarades du CRA-W avec qui les congrès RAFA sont bien plus que de simples congrès, à mes covoitureuses avec qui partir travailler devient un vrai plaisir ou encore aux membres du G10 et à nos savoureuses assemblées. Grâce à vous, je garderai de magnifiques souvenirs de ces quatre années de thèse.

Enfin, je ne serais pas en train de rédiger ces lignes aujourd'hui sans le soutien indéfectible de ma famille qui m'encourage depuis toujours. Merci à mes parents, mes beaux-parents et à Florie qui est devenue mon épouse durant ces quatre années. Merci d'avoir toujours cru en moi, de m'avoir poussé à me surpasser et pour ton extraordinaire soutien durant l'étape de rédaction. Cette thèse est également la tienne.

**Table of contents**

**ABBREVIATIONS**

ACN: Acetonitrile

AFGC: Australian Food and Grocery Council

AOAC: Association of Official Analytical Chemists

bp: base pair

C-term: Carboxyl-terminus

CI: Confidence Interval

CRM: Certified Reference Material

DNA: DeoxyriboNucleic Acid

EAACI: European Academy of Allergy and Clinical Immunology

EDTA: Ethylenediaminetetraacetic acid

ELISA: Enzyme-Linked ImmunoSorbent Assays

EW: Egg White

EY: Egg Yolk

FA: Formic Acid

FASFC: Federal Agency for the Safety of the Food Chain

G: Gibb's free energy

GFP: Green Fluorescent Protein

GRAVY: Grand Average of hydropathy

GST: Glutathione S-Transferase

GuHCl: Guanidine HydroChloride

HRMS: High Resolution Mass Spectrometry

IgA: Immunoglobulin A

IgE: Immunoglobulin E

IgG: Immunoglobulin G

ILVO: Institute for Agricultural and Fisheries Research

IPTG: Isopropyl β-D-1-ThioGalactopyranoside

IUIS: International Union of Immunological Societies

LB: Lysogeny Broth

LFD: Lateral Flow Devices

LOD: Limit Of Detection

LOQ: Limit Of Quantification

MDL: Method Detection Limit

MQL: Method Quantification Limit

MRL: Maximum Residue Level

MRM: Multiple Reaction Monitoring

MS: Mass Spectrometry

MW: Molecular Weight

m/z: mass-to-charge ratio

N-term: Amino-terminus

NVWA: The Netherlands Food and Consumer Product Safety Authority

OD: Optical Density

OFC: Oral Food Challenge

PAL: Precautionary Allergen Labelling

PAPs: Processed Animal Proteins

PCR: Polymerase Chain Reaction

PMSF: PhenylMethylSulfonyl Fluoride

Poly-H: Poly-Histidine tag

PR-10: Pathogenesis-Related protein group 10

PVDF: PolyVinyliDene Fluoride

RNA: RiboNucleic Acid

$RSD_{INT}$: Intermediate precision

$RSD_r$: Repeatability (Relative Standard Deviation)

$RSD_R$: Reproducibility (Relative Standard Deviation)

S/N: Signal to Noise ratio

SciCom: Scientific Comity (of FASFC)

SDC: Sodium DeoxyCholate

SDS-PAGE: Sodium DodecylSulfate PolyAcrylamide Gel Electrophoresis

SF: Soluble Fraction

SMPRs: Standard Method Performance Requirements

SPE: Solid Phase Extraction

SPT: Skin Prick Test

SRM: Selected Reaction Monitoring

TB: Terrific Broth

TCR: T Cell Receptor

TEAB: TetraEthylAmmonium Bicarbonate

TF: Total Fraction

TFA: TriFluoroacetic Acid

UHPLC-MS/MS: Ultra-High Performance Liquid Chromatography – tandem Mass Spectrometry

VITAL: Voluntary Incidental Trace Allergen Labelling

WAO: World Allergy Organization

WB: Western-Blot

WHO: World Health Organization

WP: Work Packages

## INTRODUCTION

### 1. Food allergy: A major public health concern

#### 1.1. Definition and epidemiology

**Food allergy** is defined by the European Academy of Allergy and Clinical Immunology (EAACI) and World Allergy Organization (WAO) as : “An adverse reaction to food mediated by an **immunological mechanism**, involving specific IgE (IgE-mediated), cell–mediated mechanisms (non IgE-mediated) or both IgE and cell-mediated mechanisms (mixed IgE and non IgE-mediated)” (Johansson *et al*, 2004). In this definition and according to Codex Alimentarius, “food” means any substances, whether they are processed, semi-processed or raw, that are intended for human consumption, and include drink, chewing gum and any substances used in the manufacture, preparation or treatment of "food" but does not include cosmetics, tobacco or substances used only as drugs (Joint FAO/WHO Food Standards Programme, 2007).

Food allergy has to be distinguished from food intolerance and coeliac disease. The three of them are classified as hypersensitive reactions to food also named adverse reactions to food. An adverse reaction that does not directly involve the immune system is considered as a food intolerance. **Celiac disease** is also considered as an adverse reaction to food but refers to an **autoimmune disorder** occurring among genetically predisposed individuals who are exposed to **gluten-containing foods** and other environmental factors (Green *et al*, 2015). Food hypersensitive reactions are summarized in the decision tree presented in Figure 1.

**Food intolerance** encompasses adverse reactions to food occurring through non-immunological reactions and embraces different mechanisms (Hayder *et al*, 2011). Lactose intolerance is a major/well known enzymatic food intolerance. Enzymatic intolerance of dietary carbohydrates can result from a variety of genetically determined enzyme deficiencies or from an illness, injury or surgery that affect the small intestine. The inability to digest a particular food compound induces a metabolic disorder. In the case of lactose intolerance, metabolic disorder results from lactase deficiency. The non-digestion of this disaccharide generates products derived from bacterial fermentation and favours the production of short chain volatile fatty acids ($C_2$-$C_6$), carbon dioxide and hydrogen gas causing symptoms like bloating, cramps and diarrhoea (Holtug *et al*, 1992).

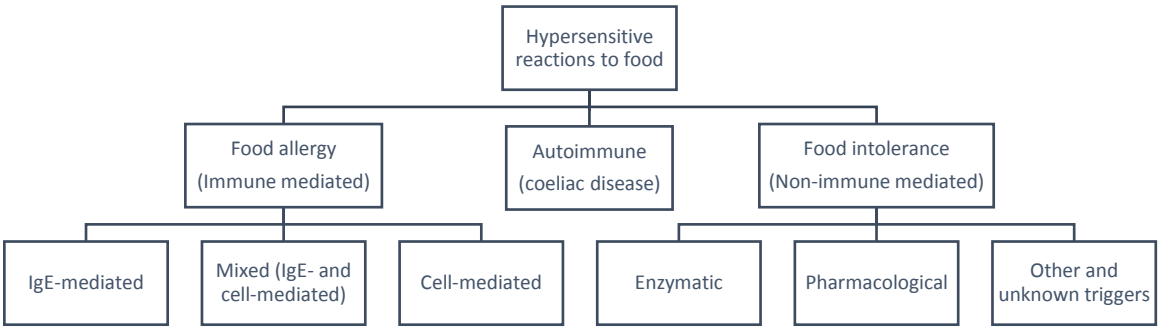



*Figure 1 - Classification of hypersensitive reactions to food adapted from EFSA Scientific Opinion on the evaluation of allergenic foods and food ingredients for labelling purposes* (EFSA, 2014)

Food allergy is triggered by **food allergens**, any food substances stimulating the production of IgE or a cellular immune response, usually a **protein**. In the context of allergen management, this term usually refers to the food or allergenic ingredients (Muraro *et al*, 2014a).

The World Health Organization/International Union of Immunological Societies (WHO/IUIS) Allergen Nomenclature Sub-Committee was created in 1986 to **standardize** the names given to the antigens (allergens) that caused IgE-mediated allergies in humans (Pomés *et al*, 2018). Approved and officially recognized allergens are now collected in an **online database** (www.allergen.org). Essential criteria for allergen acceptance are summarized in the decision chart (Pomés *et al*, 2018) (Figure 2). Several criteria such as description of the allergenic sources (including taxonomic names), information on the allergenic proteins (purification and characterization including amino acid sequences), description of the allergic human serum donors used to test IgE-binding to the candidate protein and demonstration of specific IgE-binding with five sera of relevant patients are required for the acceptance of a molecule as an allergen.

Allergens collected in the online database mainly include food allergens, but also allergens with a different route of exposure than ingestion such as injection or airway exposure. In these categories, few examples of allergens among others can be mentioned such as insect proteins transmitted through a bite or plant pollen proteins, respectively. Out of the 971 indexed allergens in December 2019, 456 entries correspond to food allergens. These food allergens are derived from 90 different species. It means that, currently, 90 food ingredients are found to be able to trigger food allergy. However, some allergens are more prevailing than others as 90 % of all food allergies are caused by "The Big 8", the 8 more prevalent allergenic ingredients including: egg, fish, milk, peanut, shellfish, soy, tree nuts, and cereals containing gluten (Koeberl *et al*, 2014).

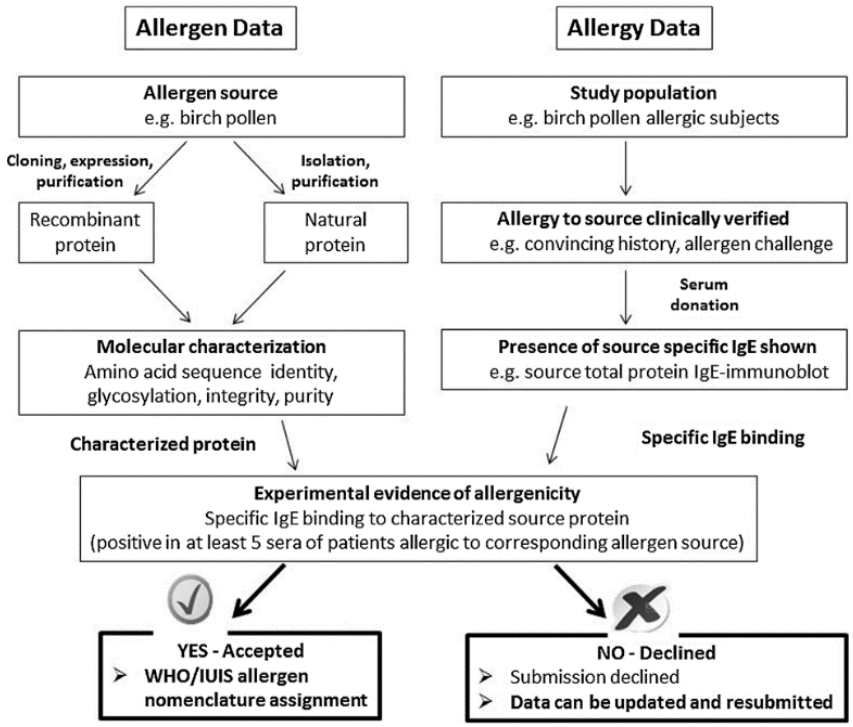



*Figure 2 - Essential criteria for allergen acceptance as determined by the WHO/IUIS Allergen Nomenclature Sub-Committee* (Pomés *et al*, 2018)

The study of **food allergy epidemiology** is challenging. Literature on food allergy prevalence is abounding, however, exposed results are relatively disparate due to numerous and variable diagnostic methodology used as well as the study of different selected sub-populations (including geographical aspects). Some studies indicate that food allergy only affects about 1 or 2 % of the population whereas other studies suggested that they can reach 10 % (Sicherer & Sampson, 2014).

Different **diagnostic methods** are proposed and used in **food allergy prevalence studies**. However, diagnostic methods differ according to implementation easiness and level of reliability. The majority of the studies are based on self-reported adverse reactions to food (Nwaru *et al*, 2014). Relying on questionnaires, they are easy to perform and can include numerous subjects. However, most of the time, studies based on self-reported adverse reactions to food over-estimate food allergy prevalence since non-immune adverse reactions to food can be confused with real food allergy. The other diagnostic methods require a clinical trial and are usually applied to selected population subgroups in which food allergy is suspected. Food allergy prevalence, based on clinical diagnostic methods, is therefore higher in such subgroups when compared to general populations. These diagnostic methods include skin prick test (SPT) (pricking the skin with a needle or pin containing a small amount of the allergen), serum-specific IgE blood tests and oral food challenges (OFC), particularly the double-blind placebo-controlled food challenges considered as the gold standard in food allergy diagnosis (Muraro *et al*, 2014b). The discrepancy between prevalence of perceived and confirmed food allergy has been reported in several studies (EFSA, 2014).

Two **meta-analyses** on food allergy prevalence were published and are powerful tools to evaluate general food allergy prevalence.

The first meta-analysis was published by Roberto J. Rona and co-workers in 2007 (Rona *et al*, 2007) and gathered data from 51 previous publications (published between January 1990 and December 2005) on food allergy prevalence. The analysis focused on 5 allergenic ingredients (milk, eggs, peanut, fish and shellfish), which were allergenic ingredients with the highest number of published reports. Results are presented separately for each one of the 5 allergenic ingredients and a pooled prevalence of food allergy to any of the 5 of them is also presented as a prevalence with a 95 % confidence interval (CI). The pooled prevalence of self-reported food allergy was 13 % for children (95% CI: 10-15 %) and 12 % for adults (95% CI: 9-14 %). The pooled prevalence of food allergy diagnosed with standardized methods and objectives diagnostic tools (SPT, serum-specific IgE blood test or OFC) was 3 % (95% CI: 2-4 %) for undistinguished children and adults (Rona *et al*, 2007). The authors insisted on caution required in estimates of prevalence based only on self-reported food allergy and highlighted the need for a collaborative study using similar methods to minimize distortions caused by both methodology and technology.

The second meta-analysis on food allergy prevalence was published by Nwaru and co-workers in 2014 (Nwaru *et al*, 2014). This meta-analysis focused on self-reported and food-challenge-defined food allergy prevalence in Europe, gathering 42 previous studies (published between January 2000 and September 2012) and including food allergy to the “Big 8” (milk, egg, wheat, soy, peanut, tree nuts, fish and shellfish). Compared to the previous meta-analysis, a higher difference was observed in self-reported food allergy between children and adults with pooled prevalence of 6.9 % (95% CI: 6.6-7.2 %) and 5.9 % (95% CI: 5.7-6.1 %) respectively. Food challenge-defined food allergy prevalence was lower with 0.9 % (95% CI: 0.8-1.1 %), without distinction between children and adults. However, milk and egg allergies were more common in younger children whereas allergy to peanut, tree nuts, fish and shellfish was more affecting older ones. A geographical difference was also observed with higher food allergy prevalence in Northern Europe when compared to Southern Europe, except for soy and peanut (Nwaru *et al*, 2014). Once again, the authors insisted on the need of studies employing standardized methodology for assessment of food allergy.

Food allergy epidemiology not only includes food allergy prevalence but also aspects as changes over time and risk factors. Many studies indicate an **increase** in food allergy prevalence over the past decades, particularly for children (Savage & Johns, 2015; Prescott & Allen, 2011). But once again, the methodology used to assess food allergy is a key point in the evaluation of time changes in food allergy prevalence. The ideal methodology would be to analyse a transversal cohort considering the same population at sequential time points with identical diagnostic methods. Unfortunately, no such studies were conducted so far. Since many food allergy prevalence studies use self-report, assessment of changes over time is limited by the potential for increased food allergy awareness in the media and other sources influencing responses over time. Most of the studies evaluated changes over time using hospital anaphylaxis admission rates or increasing health care burden as a surrogate measure for food allergy prevalence. This strategy indicated, for instance, a 2-fold increase in the United Kingdom between 1992 and 2012 (Turner *et al*, 2015). Situation in Australia is more worrisome with an increase of 13.2 % in the population between 1994 and 2005 (almost 4-fold increase). Young children (0 – 4 years) are the most affected with a 5.5-fold increase over the same period of time (Poulos *et al*, 2007).

These geographical and temporal differences in food allergy prevalence suggest that **genetic** and **environmental factors** play a role in the development of food allergy. A recent genome-wide association study identified the SERPINB gene cluster as a susceptibility locus for food allergy (Marenholz *et al*, 2017). Variants in the SERPINB gene cluster are associated with *SERPINB10* expression in leukocytes. Moreover, SERPINB genes are highly expressed in the esophagus. All identified loci are involved in immunological regulation or epithelial barrier function, emphasizing the role of both mechanisms in food allergy.

Many risk factors for food allergy have now been identified, although it is not clear what is driving the observed rise in prevalence. As in other atopic diseases, a family history of atopy is a strong risk factor (Koplin *et al*, 2013). Sex (Liu *et al*, 2010) and ethnicity (Gupta *et al*, 2011) were also reported as having an effect on food allergy. According to Liu and co-workers study (Liu *et al*, 2010), the odds of black boys having food allergy were 4.4 times higher than others in the general population. This observation needs to be carefully interpreted since the observed trend could potentially or partially also be due to different eating habits and to social and life style differences. Plethora of environmental early-life risk factors were also proposed such as microbial exposure (also known as hygiene hypothesis stating that early childhood exposure to particular microorganisms, as those from home pets, protects against allergic diseases by contributing to the development of the immune system), allergen exposure (timing and route of exposure, antacid use reducing digestion of allergen), vitamin D insufficiency (Loh & Tang, 2018), cigarette smoke and other pollutant exposure (Martino & Prescott, 2010), dietary fat (reduced consumption of omega-3-polyunsaturated fatty acids), reduced consumption of antioxidants or obesity (Sicherer & Sampson, 2014). Some of these **risk factors** need to be confirmed by complementary studies. A better understanding of risk factors will provide new avenues for **prevention** and possibly **future treatments** as discussed below.

### 1.2. Immunology of food allergy

The goal of this section is to give a rapid overview of the **principal actors in food allergy development** and allergic reactions but less deep than for a thesis in Immunology as the main objective of this thesis was to set up original analytical methods to detect and quantify four allergenic ingredients (egg, milk, peanut and hazelnut) in processed food products. As already described in the previous section, food

allergies are classified as IgE-mediated and non-IgE-mediated reactions. These two kinds of food allergies result from different mechanisms involving different cellular and molecular effectors. They will be separately described.

IgE-mediated food allergies occur when patients develop **food-specific IgE production**. Food allergy is a two phase process (Vickery *et al*, 2011). During the first step, termed allergic sensitization or **primary sensitization** (Figure 3), food allergen oral tolerance is lost and allergen-specific IgE antibodies are produced. Then, when allergen contact exposure is repeated, an IgE mediated response is induced and allergen-specific T cells are activated. This step is called the **secondary immune response** (Figure 4) and is responsible for observed symptoms of the allergic reaction (Eckl-Dorna *et al*, 2019).

Patients produce specific IgE as a result of food proteins or peptides penetrating through the gut, respiratory tract or skin. As presented in Figure 3, the antigen is processed by antigen presenting cells such as dendritic cells, macrophages or B cells which present the antigen in a major histocompatibility complex-dependent manner to T cells (Valenta *et al*, 2015). Activation of the T cell receptor (TCR) leads to cross-talk between T and B cells leading to the production of specific IgE antibodies.

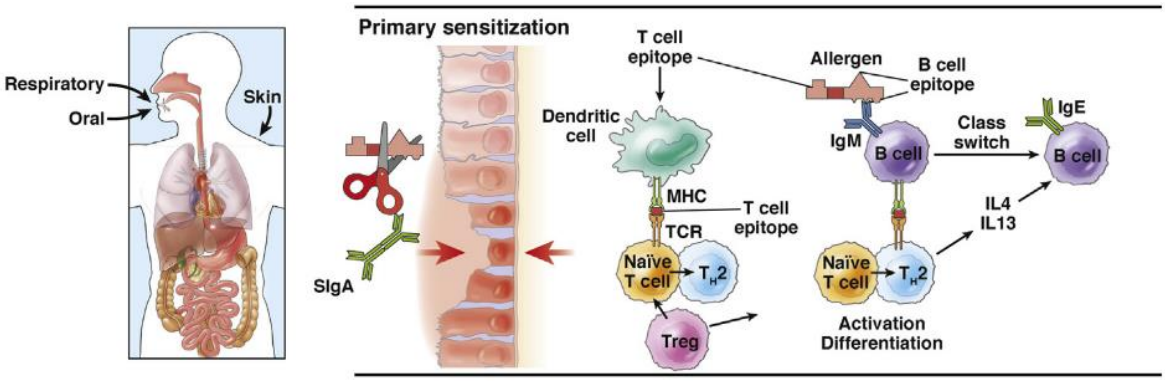



*Figure 3 – Primary sensitisation, the first step in the development of IgE-associated food allergies. Allergen contact through the gastrointestinal tract, respiratory tract or skin induces IgE production* (Valenta *et al*, 2015).

**The gastrointestinal tract** main function is to process ingested food and transform it into a form that can be absorbed, providing nutrients and energy, necessary for the growth and survival of the organism. On the other side, penetration of harmful pathogens in the body has to be prevented. Oral tolerance depends on an intact and immunologically active gastrointestinal barrier. This barrier includes physico-chemical means such as a thick mucus layer, luminal and brush border enzymes, bile salts, and the gastric acidic pH, which contribute to make antigens less immunogenic. In addition, innate immunity involving mainly natural killer cells, polymorphonuclear leukocytes, macrophages, epithelial cells, and toll-like receptors) and adaptive immunity (involving intraepithelial and lamina propria lymphocytes, Peyer's patches, IgA, and cytokines) provide an active barrier to foreign antigens (Cianferoni & Spergel, 2009).

Despite this enteric protective strategy, about 2 % of ingested food antigens are absorbed and transported throughout the body in an "immunologically intact" form (Husby *et al*, 1987). The mucosal immune system encounters therefore enormous quantities of antigen and must suppress immune reactivity to harmless food components, in other words, to develop **oral tolerance**. Antigen-presenting cells, including intestinal epithelial cells and dendritic cells, and regulatory T cells play a central role in the development of oral tolerance. The latter results from multiple mechanisms and one of the prime determinants is the dose of ingested antigen (Commins, 2015). Low doses favour the activation of regulatory T cells, components of the immune system that suppress immune responses and proliferation of conventional T helper cells controlling adaptive immunity (Corthay, 2009). Higher

doses favour the induction of clonal deletion and anergy, processes that result in the physical elimination or functional inactivation of lymphocytes, respectively (Ramsdell & Fowlkes, 1990). Food allergy occurs when oral tolerance fails to develop normally or breaks down.

During the first years of life, various components of the gut barrier and immune system are **immature**, reducing their efficiency, which plays a major role in increasing the prevalence of gastrointestinal infections and food allergy in **young children** (Sicherer & Sampson, 2010). Similarly, intrinsic factors (like genetic factors) and exogenous factors (like alcohol, anti-inflammatory drugs, pathogens or stress), which can potentially reduce the barrier function of the intestinal epithelium, are factors facilitating primary sensitisation.

Alternatively, sensitisation can be facilitated, when the gastrointestinal barrier is **bypassed** by the presentation of proteins by alternative routes, such as the respiratory tract or skin (Cianferoni & Spergel, 2009).

Oral tolerance loss due to sensitisation through the **respiratory tract** is known as pollen-food-related syndrome. Food allergy is triggered by the cross-reactivity between some pollen allergens and allergens contained in fruits. A well described example is food allergy to apple in patients sensitized to Bet v 1, the major birch pollen allergen (Olcese *et al*, 2019). Mal d 1, the major allergen of apple, and Bet v 1 are homologous proteins of the PR-10 family (pathogenesis-related protein group 10). The PR-10 proteins (16–18 kDa) defend plants against fungi and other microorganisms, and they share a common tertiary structure (Scala *et al*, 2017). Each specific domain of an antigenic molecule recognized by an antibody or a T cell receptor defines an epitope. Epitopes have approximately 15 amino acids when defined by spatial contact between antibody and epitope during binding (Frank, 2002). The high degree of sequence identity between Bet v 1 and Mal d 1 leads to IgE cross-reactivity (Fritsch *et al*, 1998). Moreover, Mal d 1 is usually well tolerated when digested, due to its instability in the presence of digestive enzymes. As described below, the allergen itself and its structure and properties have an influence on sensitisation through the gastrointestinal tract.

For a long time eczema has been known to be a major risk factor for food allergies (Abernathy-Carver *et al*, 1995). But in recent years, **skin** was found to be an important route of food allergen sensitization (Tordesillas *et al*, 2017). For example, the use in infants of diaper cream containing peanut oil or the use of facial soap containing wheat proteins were identified as risk factors for peanut and wheat allergies, respectively (Fukutomi *et al*, 2014; Lack *et al*, 2003). Epidemiologic evidence shows that an environmental exposure to peanut proteins present in house dust is a risk factor for the development of peanut allergy, particularly in those with filaggrin mutations, a protein essential to maintain the epidermis barrier (Brough *et al*, 2014). Moreover, in several mouse models, the putative role of skin sensitisation was demonstrated (Navuluri *et al*, 2006; Strid *et al*, 2004).

IgE antibodies resulting from primary sensitization circulate and bind to the IgE receptors on the surfaces of effector cells (mast cells and basophils). This step is called the **secondary response** (Figure 4). A second exposure to the allergen leads to the degranulation of these effector cells and the release of proinflammatory mediators including preformed granule products, newly synthesized arachidonic acid derived products and cytokines (He *et al*, 2013). These mediators are circulating in the bloodstream to distant tissues and have the ability to induce immediate reactions such as vasodilatation or mucous secretion as well as late-phase and chronic inflammation with influx of other inflammatory cells such as eosinophils and basophils. These phenomena are responsible for **local or systemic manifestations**, called anaphylaxis, of food allergy (Figure 4).

The most prevalent **symptoms** are **cutaneous** manifestations, including urticaria, flushing, generalized pruritus or angio-oedema (Ho *et al*, 2014). Other manifestations can affect **respiratory system** with

symptoms such as bronchoconstriction, dyspnoea, laryngeal oedema, cough, wheeze and/or hoarseness, rhinorrhoea and/or sneeze, periocular, nasal and/or oropharyngeal pruritus or mucus production (James, 2003). The **gastrointestinal system** can also be affected with throat discomfort, mouth and tongue itchiness, nausea, vomiting, abdominal cramps, and diarrhoea (Vighi *et al*, 2008). Finally, severe and potentially fatal manifestations of food allergy affect the **cardiovascular system** with symptoms including hypotension, vascular collapse or arrhythmia (Triggiani *et al*, 2008). All these manifestations of allergic reactions can be presented singly or in combination up to the **anaphylactic shock**, the most severe form that is life-threatening (Mali & Jambure, 2012).

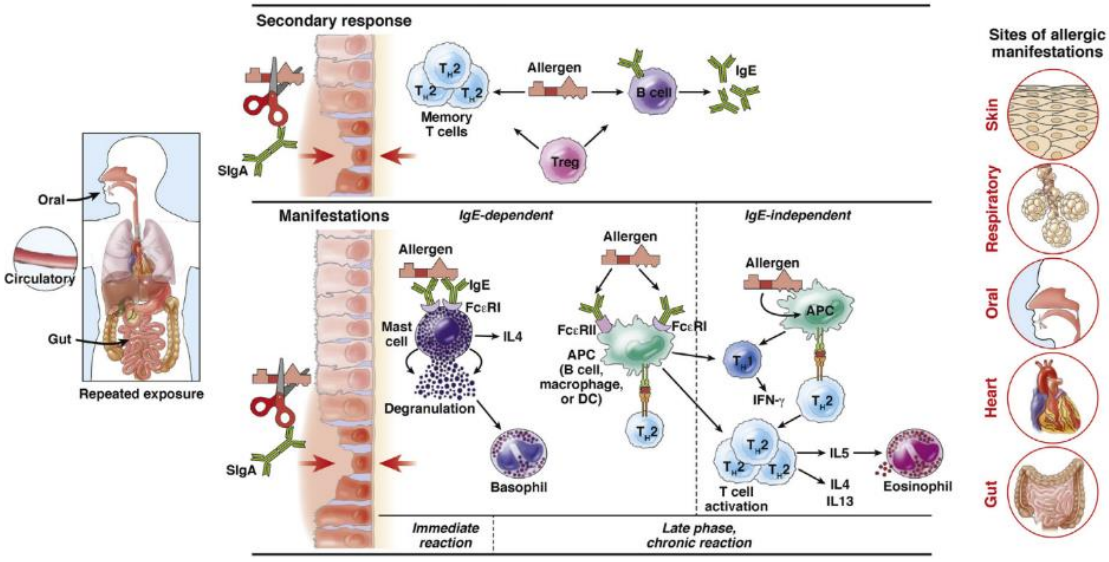



*Figure 4 – Repeated allergen contacts activate allergen-specific T cells and induce IgE-mediated responses during the secondary immune response. Degranulation of effector cells such as mast cells and basophils and release of mediators induce local or systemic manifestations/symptoms of food allergy* (Valenta *et al*, 2015).

Even if, in theory, any food protein can trigger an allergic response (Sicherer & Sampson, 2006), in practice, relatively **few protein families** account for the vast majority of allergic reactions (Sicherer & Sampson, 2010). This observation suggests that the allergen itself and its properties play a crucial role in the development of food allergy. Indeed, several shared **characteristics** were identified among major food allergens. They are usually water-soluble glycoproteins, 10 to 70 kDa in size and relatively stable to heat, acid, and protease treatments. Food processing is also believed to have an important effect on protein allergenicity (Sicherer & Sampson, 2010). For example, protein modifications such as Maillard reactions generating many different products of non-enzymatic reactions between reducing sugars and amino groups of proteins at high temperatures, are known to increase protein stability (de Oliveira *et al*, 2016). The complex structure of protein polysaccharides may cause steric hindrance preventing the digestive enzymes from reaching the binding sites. Maillard reactions are also known to induce changes in protein surface charge, hydrophobicity, tertiary structure, cross-linking, and polymerization (Lund & Ray, 2017). Thermal treatment can also lead to protein denaturation resulting in modifications of conformational epitopes (epitopes resulting from protein conformation, by opposition to linear epitopes). Hence, some patients are able to tolerate food products when they are heated but develop an allergic reaction to unheated product, as observed for some egg allergies (Ho *et al*, 2014).

Pathogenesis of **non-IgE-mediated food allergy** is less described in the literature and is still currently relatively poorly understood (Ho *et al*, 2014). However, an allergen-specific T cell-mediated process, activated by IgE-independent pathways, is strongly hypothesised (Carrard *et al*, 2015). Clinical symptoms are subacute or chronic in nature and usually displayed with isolated gastrointestinal

symptoms. Food protein-induced enterocolitis (inflammation of the digestive tract, involving enteritis of the small intestine and colitis of the colon), proctitis (inflammation of the anus and the lining of the rectum), proctocolitis (inflammation of the rectum and colon), celiac disease (long-term immune disorder caused by a reaction to gluten, that primarily affects the small intestine), dermatitis herpetiformis (chronic, intensely itchy, blistering skin manifestation), and Heiner's syndrome (pulmonary hemosiderosis) are forms of food allergy with a non-IgE-mediated immunological basis (Ho *et al*, 2014).

### 1.3. Food allergy treatment and management

In some cases, time allows to **manage** food allergy as many allergic patients naturally outgrow it over time (Renz *et al*, 2018). It has been shown that the ability to **develop tolerance** is related to the causative allergen. Egg and milk allergies are, for example, frequently outgrown whereas allergy to tree nuts or peanut tends to persist for life (Jo *et al*, 2014). The EuroPrevall birth cohort study evaluated the frequency of food allergies in 2049 children from 9 European countries from birth until 2 years of age (Schoemaker *et al*, 2015). The authors showed that 1 year after egg and milk allergy diagnosis, half and more than half of the allergic children developed tolerance, respectively. By contrast, the Australian study HealthNuts showed that only 22 % of children diagnosed with peanut allergy at 1 year of age, developed tolerance when they were 4-year-old (Peters *et al*, 2015).

Allergen immunotherapy or **desensitization** is already applied to treat venom allergy and allergic rhinitis and conjunctivitis caused by inhalant allergens (Arshad, 2016). This strategy is currently under development to treat food allergy. The objective is to induce tolerance to specific allergens and is based on a two-step process. Patients are first desensitized by regular administration of the allergen. The goal is to **reduce or eliminate responses of effector cells** involved in the specific immune response. Then, tolerance, in which the nonreactive state permanently remains, is established. Once a state of tolerance is achieved, the patient can consume the food infrequently without eliciting a response (Beyer, 2012). Different administration routes are explored, including oral, sublingual, subcutaneous and epicutaneous routes (De Silva *et al*, 2014).

While **promising**, allergen immunotherapy applied to food allergy is not successfully accomplished, yet. Indeed, not all studies have shown a benefit of this approach and the **risk of adverse reactions** such as anaphylaxis has to be considered (Carrard *et al*, 2015). Moreover, induced tolerance may only be **transient**. Some patients were found to lose their tolerance after discontinuation of a maintenance dose of the allergen (Mota *et al*, 2018). Consequently, the current approach to food allergy management mainly relies on **allergen avoidance** and preparation to promptly treat allergic reactions. Administration of adrenaline/epinephrine by intramuscular injection is the standard **treatment** for systemic reactions (McLean-Tooke *et al*, 2003). Adrenaline can rapidly (within minutes) reverse oedema, urticaria, bronchospasm, hypotension and gastrointestinal symptoms. Administration timing is a crucial factor notably in preventing death from anaphylactic shock (Yu *et al*, 2016). The more rapid the injection after allergen exposure, the more efficient the treatment. Other pharmaceuticals, such as anti-histaminic drug, are also used to treat localized food allergy symptoms (Randall & Hawkins, 2018).

In the absence of recognised and accepted treatments, food allergy management is based on **dietary management** with the exclusion of the offending allergenic food. To efficiently implement this strategy, the allergic consumer has to be able to easily identify the presence of the allergenic ingredient in a given food product. In this regard, many jurisdictions have introduced **labelling requirements** for

substances causing allergies and intolerance reactions in sensitive individuals (Popping & Diaz-Amigo, 2018).

The first **European allergen-labelling directive** was passed in 2003. The directive No. 2003/89/EC amended the directive No. 2000/13/EC with regards to the indication of the ingredients present in foodstuffs. Annex IIIa was added to the directive No. 2000/13/ EC to include a list of allergenic ingredients and substances causing intolerances that required labelling. This first directive included a list of 12 allergens and products thereof without any exemptions. However, some derived materials are known not to contain food-allergy-triggering proteins. Food products containing such materials are therefore harmless for allergic consumers. The restrictiveness of this first directive adversely affected the economy of the food industry and the quality of life of affected individuals alike.

In 2005, a temporary exception list was published in the directive No. 2005/26/EC. A series of materials derived from allergenic ingredients, such as wheat-based glucose syrups, nuts used in distillates for spirits or fully refined soybean oil were temporary exempted to the amended labelling directive. The final list of permitted exemptions was published in 2007 in the directive No. 2007/68/EC (9), amending aforementioned Annex IIIa of the directive No. 2000/13.

In 2011, after multiple corrections and amendments and the addition of two new allergenic ingredients (lupine and molluscs), the directive No. 2000/13/EC was repealed by the **Consumer Information Regulation (EC) 1169/2011**. This new regulation includes the list of allergenic ingredients and substances causing intolerances that required labelling and permitted exemptions in Annex II (detailed in Table 1).

This regulation was positively considered by consumer and patient advocacy groups. However, this regulation is **limited to food ingredients**, meaning foods or substances that are voluntarily incorporated to the food products, that are part of the recipe. Consequently, this regulation does not address a **major problem** in food allergy management, **cross-contamination**. A cross-contamination corresponds to the inadvertent introduction of an allergen into a product that would not intentionally contain that allergen as an ingredient, and therefore **not labelled as containing the allergen**. Cross-contamination may result when multiple foods are produced in the same facility or on the same processing line, through the misuse of rework, as the result of ineffective cleaning, or may result from customary methods of growing and harvesting crops, as well as from the use of shared storage, transportation, or production equipment (Figure 5). Numerous allergic reactions were identified to be triggered by cross-contaminated food products (Gendel *et al*, 2008). Much cross-contamination can be avoided by controlling the production environment.

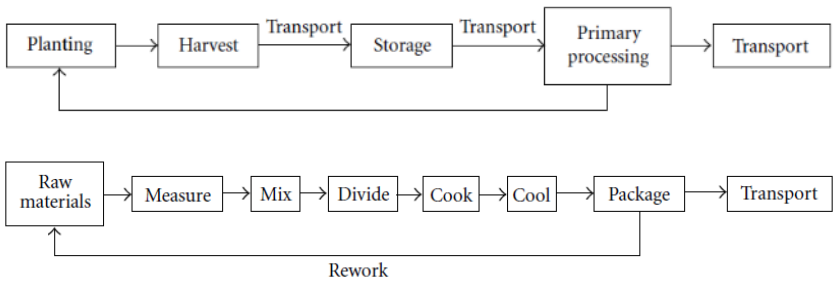



*Figure 5 – Schematic of stages involved in food production and manufacturing process. Cross-contaminations can occur at any point in the process such as the use of contaminated seeds in the next planting season or through rework, and the incorporation of pre-worked packaged food into new production batches* (Alvarez & Boye, 2012)

*Table 1 - List of allergenic ingredients and substances causing intolerances that require labelling and permitted exemptions from Annex II of Consumer Information Regulation (EC) 1169/2011*

1. Cereals containing gluten, namely: wheat, rye, barley, oats, spelt, kamut or their hybridised strains, and products thereof, except:
   (a) wheat based glucose syrups including dextrose[1]
   (b) wheat based maltodextrins[1]
   (c) glucose syrups based on barley
   (d) cereals used for making alcoholic distillates including ethyl alcohol of agricultural origin
2. Crustaceans and products thereof
3. Eggs and products thereof
4. Fish and products thereof, except:
   (a) fish gelatine used as carrier for vitamin or carotenoid preparations
   (b) fish gelatine or Isinglass used as fining agent in beer and wine
5. Peanuts and products thereof
6. Soybeans and products thereof, except:
   (a) fully refined soybean oil and fat[1]
   (b) natural mixed tocopherols (E306), natural D-alpha tocopherol, natural D-alpha tocopherol acetate, and natural D-alpha tocopherol succinate from soybean sources
   (c) vegetable oils derived phytosterols and phytosterol esters from soybean sources
   (d) plant stanol ester produced from vegetable oil sterols from soybean sources
7. Milk and products thereof (including lactose), except:
   (a) whey used for making alcoholic distillates including ethyl alcohol of agricultural origin
   (b) lactitol
8. Nuts, namely: almonds (*Amygdalus communis L.*), hazelnuts (*Corylus avellana*), walnuts (*Juglans regia*), cashews (*Anacardium occidentale*), pecan nuts (*Carya illinoinensis* (Wangenh.) K. Koch), Brazil nuts (*Bertholletia excelsa*), pistachio nuts (*Pistacia vera*), macadamia or Queensland nuts (*Macadamia ternifolia*), and products thereof, except:
   (a) nuts used for making alcoholic distillates including ethyl alcohol of agricultural origin
9. Celery and products thereof
10. Mustard and products thereof
11. Sesame seeds and products thereof
12. Sulphur dioxide and sulphites at concentrations of more than 10 mg/kg or 10 mg/L in terms of the total $SO_2$ which are to be calculated for products as proposed ready for consumption or as reconstituted according to the instructions of the manufacturers
13. Lupine and products thereof
14. Molluscs and products thereof

[1] And the products thereof, in so far as the process that they have undergone is not likely to increase the level of allergenicity assessed by the Authority for the relevant product from which they originated.

Given the multiple potential sources of cross-contamination and the risk posed to allergic individuals by even very low residual amounts of allergens, many food manufacturers provide advice related to the potential of unintentional contamination with allergens during manufacture in the form of **precautionary allergen labelling (PAL)**, also known as "**may contain**" statements. However, except in South Africa and Switzerland, the use of PAL is not regulated by legislation, and it is suspected that in many cases, a formal risk assessment is not performed to guide the use of PAL (Allen *et al*, 2014). The widespread use of PAL has even negative effects on allergic consumers as the use of different wording on PAL statements is **confusing** and can lead to miss the proper ingredient-based allergy warning. A survey even indicated that more than 40 % of young American adults with food allergy simply ignore PAL (Sampson *et al*, 2006). The loss of trust in PAL thus reduces the ability of consumers with food allergies to make informed choices leading to reduced avoidance, reduced quality of life and increased risk-taking by consumers ignoring PAL.

All contributing stakeholders (including clinicians, patients, food industry and regulators) agree that PAL must reflect the actual risk. The use of PAL should be based on a **risk assessment** and should indicate the possible, unintended presence of an allergen in a consumed portion of a food product at or above any proposed action level (DunnGalvin *et al*, 2015). This approach involves numerous concepts and requires the collaboration of all stakeholders.

The **Voluntary Incidental Trace Allergen Labelling (VITAL)** initiative developed by the Australian food industry's Allergen Bureau represented a first attempt to introduce a formal and transparent basis for **risk assessment** by manufacturers in the application of PAL (Koplin *et al*, 2010). The objective of this VITAL initiative is to exclusively identify the need of PAL using a validated risk assessment tool.

VITAL approach is based on the evaluation of the putative presence of allergen residues arising from cross contact and the calculation of the resulting final allergen content. This calculation integrates a product information form of all raw materials to evaluate the risk of allergen contamination of each ingredient. Production lines and manufacturing environment are also reviewed to identify potential cross-contact during the manufacturing process. The VITAL calculator determines the final allergen content which is compared to **Reference Doses** defined for major allergenic foods. These Reference Doses are based on **clinical data** related to the minimal allergen protein quantity/amount triggering an allergic reaction. Since these doses might vary for different patients, the selected thresholds are deemed to be tolerated by 99 % ($ED_{01}$) of the allergic population. Until the end of 2019, thresholds for the less common foods such as celery or cashew with sparser clinical data were deemed to be tolerated by 95 % ($ED_{05}$) of the allergic population. However, the VITAL expert panel now recommends the adoption of $ED_{01}$ values as the Reference Doses for **VITAL 3.0** (Hawkes, 2019). Reference Doses for the **14 allergenic ingredients** considered in VITAL 3.0 program are listed in Table 2.

Reference Doses correspond to protein amounts whereas the VITAL allergen content calculator evaluates allergen concentrations. To be able to compare these different values, the "**serving size**" concept was introduced. It corresponds to the maximum amount of a food eaten in a typical eating occasion. For this conversion, the VITAL organization recommends to apply serving size principles of the Australian Food and Grocery Council (AFGC).

*Table 2 – Comparison of Allergen Reference Doses between VITAL 3.0, provisional doses proposed by SciCom of FASFC in Belgium in 2017 and preliminary doses proposed by NVWA in The Netherlands in 2016. The doses are corresponding to the minimal allergen protein quantity (expressed in mg) triggering an allergic reaction in 5 % (Belgium) or 1 % (VITAL and The Netherlands) of the allergic population.*

| Allergen | Reference Dose (mg protein) | | |
|---|---|---|---|
| | VITAL 3.0 | SciCom, FASFC (BE) | NVWA (NL) |
| Egg | 0.2 | 0.3 | 0.0043 |
| Hazelnut (and Tree Nut default) | 0.1 | 0.5 | 0.011 |
| Lupin | 2.6 | 4.5 | 0.83 |
| Milk | 0.2 | 1.2 | 0.016 |
| Mustard | 0.05 | 0.1 | 0.022 |
| Peanut | 0.2 | 1.1 | 0.015 |
| Sesame | 0.1 | 0.4 | 0.10 |
| Shrimp | 25 | 12.1 | 3.7 |
| Soy (milk + flour) | 0.5 | 2.9 | 0.078 |
| Wheat | 0.7 | 1.3 | 0.14 |
| Cashew (and Pistachio) | 0.05 | 0.6 | 1.4 |
| Walnut (and Pecan) | 0.03 | 0.5 | / |
| Celery | 0.05 | / | / |
| Fish (finfish) | 1.3 | / | / |

Using VITAL calculator, serving size and Reference Doses as references, decisions on the use of PAL for allergen cross contamination can be made. There are two possible situations defined as **action levels**. When the calculated allergen content is below the threshold, it corresponds to Action Level 1 where a precautionary cross contact statement is not required. In other words, the food is considered as safe to eat by the allergic consumers. On the contrary, when the calculated allergen content is above the threshold, a precautionary cross contact statement is required, corresponding to Action Level 2.

With the peanut example below (peanut Reference Dose = 0.2 mg), we see that serving size directly affects Action Levels:

$$\begin{array}{cc} 5\ g\ serving\ size & 50\ g\ serving\ size \\ Action\ level\ 1\colon\ < 40\ ppm & Action\ level\ 1\colon\ < 4\ ppm \\ Action\ level\ 2\colon\ \geq 40\ ppm & Action\ level\ 2\colon\ \geq 4\ ppm \end{array}$$

Importantly, under the VITAL approach, only a **single distinct precautionary statement** "may be present" is to be used, in order to avoid the confusion which has resulted from multiple phrases used in PAL.

However, VITAL proposed thresholds have no legal status. Based on the same methodology, different countries proposed allergen reference doses to manage the risk that may arise from the presence of allergens in foods. The problem is that these thresholds differ greatly from country to country, even in Europe.

In Belgium, the scientific comity of the Federal Agency for the Safety of the Food Chain (SciCom of FASFC) proposed, in 2017, provisional allergen reference doses based on reported scientific literature data (not only VITAL). These reference doses are mainly based on the lower limit of the 95 %

confidence interval of the dose that elicits an allergic reaction in 5 % of the susceptible population ($ED_{05}$) (SciCom, 2017).

The Netherlands Food and Consumer Product Safety Authority (NVWA) proposed, in 2016, preliminary reference doses relying on published data in the literature (NVWA, 2016). These reference doses are based on the lowest $ED_{01}$ value of the models used (Remington, 2013).

Belgian and Dutch reference doses are exposed in Table 2, highlighting the **differences between countries**. This situation is complicated for food industrials and products exportation. A **harmonized** regulatory framework for managing hidden allergens or action thresholds, at the European level is therefore strongly requested by food producers and control laboratories.

**Food allergen analysis** is an essential tool in the development of a **risk-based approach to allergen management**. Food allergen analysis is necessary for the testing of a material or a surface to detect, identify and quantify the presence of food allergens. Test results can provide assurance and verification of critical controls within a comprehensive allergen management plan and assist the implementation of quantitative risk assessment.

The following section is dedicated to the different aspects of different methods developed for food allergen analysis and will discuss the encountered challenges, advantages and limitations. A particular attention will be devoted to possibilities of allergen identification and quantification by mass spectrometry, as the development of such methods for routine analytical laboratories was the primary objective of this thesis.

## 2. Food allergen analysis: An essential tool towards safe food products

### 2.1. Challenges in food allergen analysis

The analysis of food allergens is a challenging task. Traditionally, in residues analysis, a defined compound or a derived molecule is targeted by the analysis method and a **maximum residue level** (MRL) is defined by the legislation. This is the case for many types of contaminants analysed in food including veterinary drugs, pesticides, toxins (mycotoxins, phycotoxins or plant toxins), food dyes or heavy metals. In food allergen analysis, the European legislation states:

*Regulation (EU) No 1169/2011 of the European Parliament*

*"When used in the production of foods and still present therein, certain ingredients or other substances or products (such as processing aids) can cause allergies or intolerances in some people, and some of those allergies or intolerances constitute a danger to the health of those concerned. It is important that information on the presence of food additives, processing aids and other substances or products with a scientifically proven allergenic or intolerance effect should be given to enable consumers, particularly those suffering from a food allergy or intolerance, to make informed choices which are safe for them."*

In other words, the presence of the allergenic ingredient in a given food product has to be communicated to the consumer, through food labelling for prepacked food. Consequently, the objective of food allergen analysis is to detect and eventually quantify the presence of the allergenic ingredient in a given sample, regardless of the chosen analyte.

In general, it is preferable to employ **analytical methods** that target the hazard; hence, methods able to determine the presence of allergenic proteins per se should be used in preference to others (Muraro *et al*, 2014a). However, European legislation requires the labelling of the ingredient and not the

allergenic proteins. Therefore, the majority of current methods do not target allergens but are, instead, based on the **detection of indicator proteins, peptides or nucleic acids**. Indeed, the three main methods employed for food allergen analysis are DNA-based methods, immunological methods and mass spectrometry-based methods. Principles, advantages and drawbacks of these three techniques are developed in the following sections with a particular attention and focus on mass spectrometry-based methods and their potential quantitative application justified by the objectives of our work.

Even if multiple components are targeted for food allergen analysis, a crucial aspect in food allergen management is the use of harmonized **reporting units**. As developed in the previous section, there are currently no harmonized legal thresholds for allergen contamination. Analysis laboratories therefore consider VITAL achievements or country specific proposed thresholds as in Belgium. Defined Reference Doses are based on clinical data and are related to the minimal allergen protein quantity triggering an allergic reaction. These Reference Doses are expressed in **total protein amount from the allergenic ingredient**. Using the "Serving Size" concept, this protein amount is transformed in **concentration expressed in ppm** or mg total protein from the allergenic ingredient per kg of food matrix.

Another main issue in food allergen analysis is the complexity and diversity of food matrices that can be encountered and their associated **matrix effects** impacting/affecting/altering the performance of the different analytical methods. To develop further the importance of the interfering effect of the matrices, one can mention that the analyte has to be detected at a trace level, in the order of ppm (mg per matrix kg), in an environment that can contain multiple protein and DNA sources and other **interfering components**. Moreover, food products are rarely consumed in their raw form. Applied food **processes** (such as cooking, pasteurization, sterilization, drying, evaporation, distillation, chilling, freezing, etc.) may have an impact as they can physically or chemically modify the targeted molecules that could affect their **extractability** and **detectability** (Walker *et al*, 2016).

### 2.2. DNA-based methods

DNA-based methods rely on the detection of specific DNA sequences coding for the allergenic protein, or other **DNA sequences specific for the allergenic food of interest**. Polymerase Chain Reaction (PCR) is the dominant DNA-based technique used for the detection of allergenic ingredients (Nollet & van Hengel, 2010).

General PCR analysis consists of three main steps (Figure 6):

- DNA extraction from food samples and purification involving cell lysis, DNA solubilisation and chemical or enzymatic methods to remove macromolecules, lipids, RNA, or proteins.
- Amplification of a specific DNA sequence through a repetition of thermocycles alternating (1) DNA melting (denaturation of the double-stranded DNA template by breaking the hydrogen bonds between complementary bases, yielding two single-stranded DNA molecules), (2) primers annealing (binding of DNA primers to DNA template) and (3) DNA polymerization (synthesis of a new DNA strand complementary to the DNA template strand by DNA polymerase).
- Detection of the generated PCR products.

Several methods exist to detect the PCR products.

Agarose gel electrophoresis was historically used to detect the PCR products according to size, in the "classic PCR" technique. However, the specificity of this technique merely relies on the primer selection, and the electrophoretic separation of DNA bands. Moreover, this technique is generally not

suitable for quantitative analysis because the amount of PCR products is formed up to a plateau of maximum product accumulation, independently of the starting amount of DNA (Holzhauser & Röder, 2015).

For an advanced sequence verification of the PCR product, sequence-specific DNA probes are additionally applied. These probes are applied after the amplification in the case of post-PCR detection or during the amplification in real-time PCR.

An example of post-PCR detection is PCR-ELISA where the amplified DNA is detected by an ELISA-like technique that involves binding of the PCR products to the surface of a microplate and its subsequent detection by enzyme conjugated sequence-specific probes (see section 2.3 on page 29 for detailed description of the ELISA technique). Holzhauser and co-workers developed a PCR-ELISA method to detect hazelnut in processed food products at a level below 0.001 % (w/w) (Holzhauser *et al*, 2002).

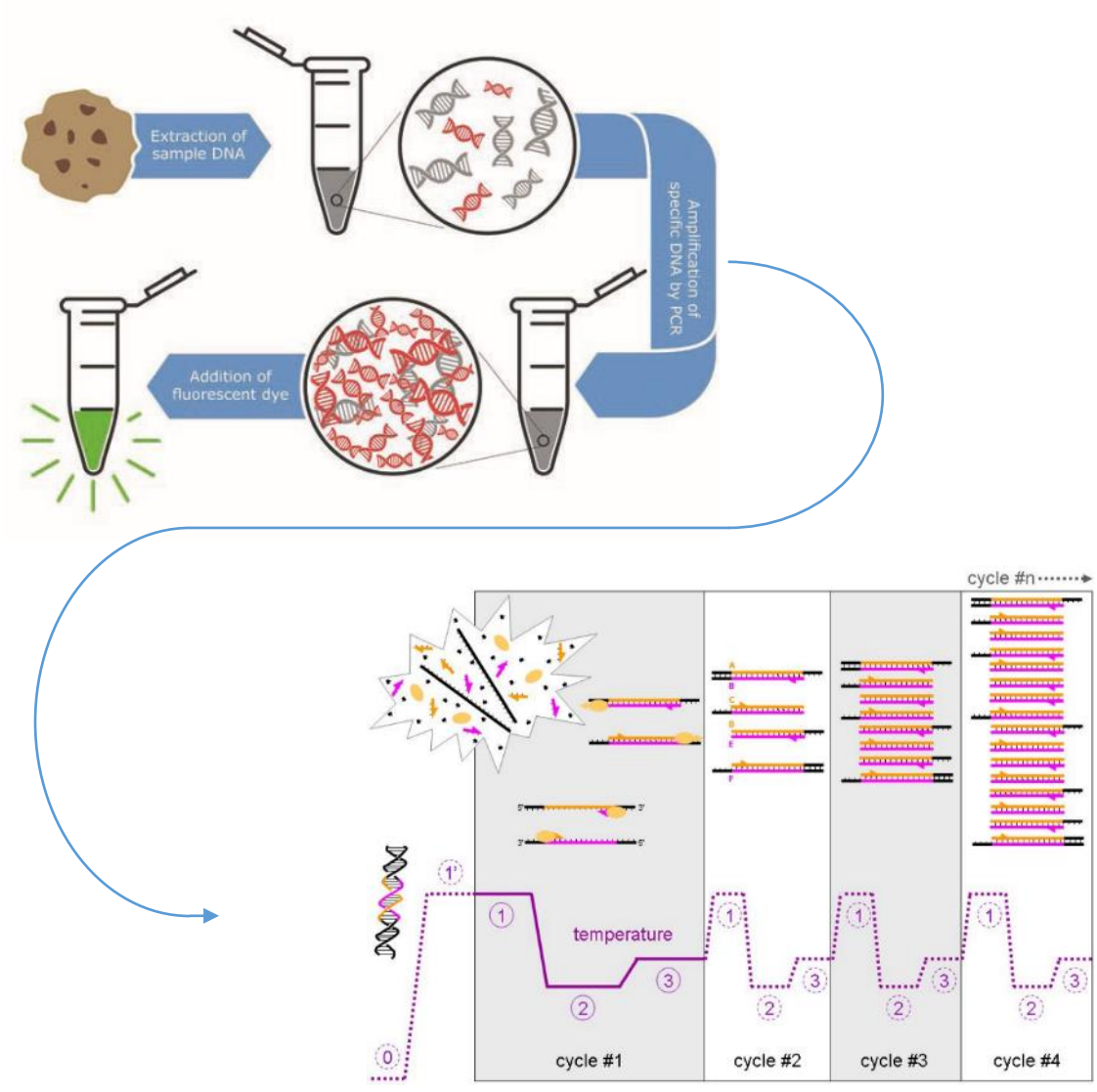



*Figure 6 - A simplified scheme of PCR analysis from Romer Labs website* (Brunner, 2016) *and work of Issa* (Issa, 2013). *DNA specific amplification is performed by the repetition of thermocycles comprising three steps: (1) DNA melting, (2) primers annealing and (3) DNA polymerization.*

**Real-time PCR** is the most widely applied PCR technique to detect food allergens (Linacero *et al*, 2016) with commercial kits available for the detection of several food allergens including celery, gluten, lupine, milk, peanut, sesame, soya and tree nuts (almond, brazil nut, cashew, hazelnut, macadamia,

pecan, pine nut, pistachio and walnut) (Poms *et al*, 2004; Linacero *et al*, 2020; Słowianek & Majak, 2011). Using specific oligonucleotide probes carrying a **reporter dye**, the amount of amplified DNA is recorded within each amplification cycle. The gain of detectable fluorescence is directly proportional to the increased amount of amplified DNA. On the basis of suitable calibrants, the amount of target DNA or food can be **quantified**.

The first and critical step in the development of a PCR analysis method is the selection of the targeted DNA sequence. In principle, any nucleotide sequence from the target species that can be specifically discriminated from nucleic acid sequences of other food components is suitable for specific detection by PCR (Holzhauser & Röder, 2015). For example, Röder and co-workers were able to specifically detect almond allergen and discriminate it from other phylogenetically closely related foods of the *Prunoideae* family, such as peach, apricot, and plum (Röder *et al*, 2011). It was achieved on the basis of only two differing nucleotides within the amplified gene sequence coding for a non-specific lipid transfer protein.

This characteristic gives a **high specificity** to DNA-based food allergen detection methods. However, targeting DNA can be a severe drawback in the case of milk and egg analysis. Both ingredients are recognized as allergenic foods. The main issue is that milk and eggs share identical DNA sequences with beef and chicken, respectively. Therefore, they cannot be differentiated from their respective meat which is not considered as allergenic food (Flanagan, 2014). Moreover, the abundance of DNA in milk and egg is very low, which can lead to sensitivity issues (Köppel *et al*, 2010). DNA-based methods are therefore **not suitable for the detection of two major allergenic food ingredients, egg and milk**.

In the case of **celery**, the opposite situation is encountered. According to relatively recent literature, **no specific antibody** has successfully been developed for the immunological detection methods. All attempts to produce reliable antibodies have failed due to the cross-reactivity of the antibodies raised against celery proteins and proteins from other plants such as parsley, carrot, coriander or fennel (Flanagan, 2014). These aspects of antibody cross-reactivity are more extensively described in the next section dedicated to immunological methods. To overcome these limitations, PCR is currently used to specifically detect celery (Hupfer *et al*, 2007).

In terms of sensitivity, PCR for the detection of food allergens has a **method detection limit of less than 10 mg/kg**. This value is expressed in mg total ingredient per kg matrix and is in the range of VITAL requirements (Kirsch *et al*, 2009). Theoretically, under optimal conditions, one single copy of the targeted DNA sequence per PCR reaction should be sufficient for a successful detection (Holzhauser & Röder, 2015). However, in practice, multiple factors, detailed in the next paragraph, are known to influence PCR analysis performance. In addition, the choice of the targeted nucleotide sequence can influence the sensitivity of this method. A 10 to 100 times more sensitive detection may be achieved when targeting a multi-copy sequences (a repeated sequence) from organelles such as mitochondria or chloroplasts (Holzhauser & Röder, 2015). However, higher sensitivity may be impaired by lower specificity if the multi-copy sequences are highly conserved throughout evolution. In any case, the specificity needs to be verified experimentally with the analysis of related species.

**DNA is generally more stable than proteins** and thus should be more suitable for the use of harsher extraction conditions (Walker, 2018). However, matrix impurities can significantly affect/lower the efficiency of the detection and the presence of these impurities have been described as being more critical to DNA than to protein analysis (Flanagan, 2014). Components such as fats, polysaccharides or minerals can also inhibit polymerization reaction (Schrader *et al*, 2012). The impact of food processing on protein is detailed in sections dedicated to immunological and mass spectrometry-based methods.

Even if often described as more stable than protein, it was demonstrated that DNA detection by real-time PCR can also be impacted by food processing (Platteau *et al*, 2011).

However, DNA-based methods offer the **possibility for simultaneous detection of multiple allergenic** food ingredients (multiplexed analyses). DNA extraction is independent of the target and any target gene can, in principle, be amplified from one single DNA preparation. In a method published by Ehlert and co-workers, up to 10 allergenic food ingredients (peanut, cashew, almond, brazil nut, hazelnut, pecan, pistachio, macadamia nut, walnut and sesame) were simultaneously detected in different matrices (chocolate, cookie and pesto) at a 5 ppm contamination level (expressed here in mg total ingredient per matrix kg) (Ehlert *et al*, 2009).

**Quantitative methods** are possible with DNA-based methods but the main problem is the absence of **certified reference material (CRM)**. The European Standard EN 15634-1:2009 relative to the detection of allergenic ingredients in foodstuffs by molecular biology methods based on DNA analysis establishes that the method detection limit (MDL) and the method quantification limit (MQL) must be expressed as the number of copies of DNA equivalent to a total quantity of the allergenic ingredient per kilogram of food (mg/kg). The equivalence should be based on CRM, but these are not available yet. Reference materials (not certified) developed by different producers are commercially available for most major food allergens, but the results obtained with different quantification kits may not be comparable (EFSA, 2014).

Considering practical aspects, DNA-based PCR methods require a dedicated laboratory space and trained personnel. The required financial investment can be relatively important for multiplexing quantitative instruments and complete analysis, from sample reception to final result, can be performed within a working day (Brežná *et al*, 2006).

In conclusion, DNA-based methods display a high sensitivity and specificity for the analysis of food allergens with possibilities of multiplexing and quantification. However, their use remains questionable since the targeted analyte, **DNA, is not directly responsible for food allergy health risk**. The absence of allergenic ingredient specific DNA does not necessarily mean the absence of risk for sensitized consumer. Appropriate detection methods must be carefully considered depending on sample type and analysis objective.

### 2.3. Immunological methods

Immunological methods are the most widely used techniques by the food industry to detect the presence of food allergens (Picó *et al*, 2012). These **protein-based methods** rely on the utilization of antibodies, raised to specific allergenic proteins or peptides which serve as markers for the detection of the allergenic food. Allergen detection is based on the quantitative measurement of the **specific binding of an antigen**, the allergenic protein or protein sequence, **to an antibody** (Flanagan, 2014). Given their lower production costs, polyclonal antibodies are typically used for detection kit manufacturing (Picó *et al*, 2012). Polyclonal antibodies recognize multiple epitopes on the protein and are opposed to monoclonal antibodies, with higher production costs, that are suited to the recognition of one epitope along the allergen moiety. Immunological methods can adopt different forms including enzyme-linked immunosorbent assays (ELISA), lateral flow devices (LFD) and dipsticks.

The detection of allergens by **ELISA** is mainly performed with two techniques. One is the so called 'sandwich' ELISA technique and the other is the 'competitive' format (Figure 7). The format used depends on the size of the target molecules. In the case of most allergen ELISAs, the sandwich format can be used to quantify intact allergenic proteins or polypeptides. However, the competitive format is

mandatory in certain circumstances, for example, for the analysis of small peptides which contain only one binding site for the antibody (Flanagan, 2014).

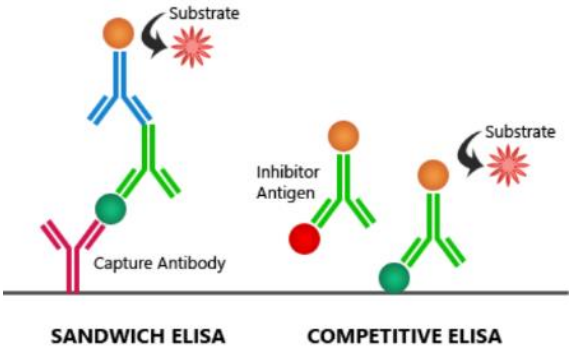



*Figure 7 – The two ELISA formats, sandwich and competitive, used for food allergen detection* (Budhathoki, 2018)

In the **sandwich ELISA format**, a capture IgG antibody (hereafter just called antibody) is immobilized onto the surface of a solid phase, typically a polystyrene microtiter plate. The extracted sample is loaded in the microwell and any allergenic protein, specific to the capture antibody, will bind to it. After a certain incubation time, the microwell is washed to remove unbound components. Then, two detection methods are possible. In the case of direct sandwich ELISA, an enzyme-conjugated antibody (for example, conjugated to horseradish peroxidase or alkaline phosphatase) or a secondary antibody is added and directly reacts with the capture antibody–antigen complex. In the case of an indirect sandwich ELISA (as shown in Figure 7), the secondary antibody is not labelled with an enzyme. A third antibody, conjugated with an enzyme, is used to detect the capture antibody–antigen-secondary antibody complex. This third antibody is specific to the secondary antibody (for example, if the secondary antibody is a goat antibody, the third antibody will be an anti-goat antibody). After a further washing step, a substrate is added and the enzyme bound to the secondary (direct) or third (indirect) antibody converts the colourless substrate into a coloured product, indicating the presence of the capture antibody–antigen-secondary antibody (direct) or capture antibody–antigen-secondary antibody-third antibody (indirect) complex. The optical density of the solution in a spectrophotometer at a particular wavelength is thus proportional to the abundance of the targeted analyte. The amount of a particular component in an unknown sample can be determined using a calibration curve performed with a suitable standard (Herman *et al*, 2008).

In the **competitive ELISA format**, the targeted compound is immobilized on the surface of the microwell. The extracted allergen is loaded in the microwell as well as a fixed amount of enzyme-conjugated antibody. After a certain incubation time, the solution is discarded, the wells are washed and the appropriate substrate is added. The enzyme-conjugated antibody, bound to the target on the surface of the microwell, catalyses the enzymatic reaction. The more target in the solution, the less enzyme-conjugated antibody can bind to the targeted compound immobilized on the surface of the microwell. The colour development is thus inversely proportional to the amount of analyte of interest. An indirect format of competitive ELISA is also possible when no enzyme-conjugated antibody specific for the analyte is available. In this case an enzyme-labelled species-specific antibody is used to detect the antigen-antibody complex (Clemente *et al*, 2004).

Lateral flow devices and dipsticks are simplified versions of ELISA. In LFDs, the sample flows along a polyvinylidene difluoride membrane by capillarity to reach a line where the antibody has been adsorbed, giving rise to a coloured antigen-antibody complex (Rudolf *et al*, 2012). Dipsticks are based on the same principle but do not have a mobile phase moving up along the strip. **LFDs and dipsticks**

**are mainly used for on-site screening** since they are inexpensive, quick, portable and easy to use. However, the obtained information is only **qualitative** and they suffer from the same limitations described below for immunological methods (EFSA, 2014).

In the food industry, because of its **high level of sensitivity**, **ease of use** and **limited cost**, ELISA methods have become the **standard method** for the qualitative detection and semi-quantitative determination of specific proteins from allergenic sources (Baumert, 2014). The MDL of food allergen ELISA methods is typically in the range of ppm (0.1 to 1.5 ppm depending on the allergenic ingredient and food matrix (EFSA, 2014)), which is **in accordance to current VITAL Reference Doses**.

However, immunological methods suffer from several **limitations** that we will now rapidly discuss. The sensitivity and specificity of the method is highly dependent on the biophysical and chemical properties (e.g., solubility, structure, conformation, and chemical alterations) of targeted allergens (Sharma *et al*, 2017). In addition, **food processing** may cause allergen conformational changes, denaturation, aggregation, chemical modifications of epitopes, or interactions with food matrix components. These changes affect protein extractability and antibody recognition of allergenic proteins. Heat processing induces phenomena such as protein aggregation or chemical reaction (i.e. Maillard reactions) rendering proteins less soluble (Albillos *et al*, 2011). Non-thermal processing, such as hydrolysis, can also alter the epitope-binding region of target proteins, reducing the antibody interaction necessary for accurate detection and quantification. In this case, the food cannot be considered to be free of allergenicity because partially hydrolysed proteins may still retain allergenic potential (Picó *et al*, 2012). Moreover, **cross-reactivity** with similar proteins from different sources can also affect the specificity of the method (Koeberl *et al*, 2018; Costa *et al*, 2016a) and lead to **false-positive results**. ELISA allows for the measurement of only one analyte at a time in a given sample. As a result, when testing food for multiple food allergens and gluten, it becomes necessary to perform multiple analytical tests, a time-consuming and expensive process (Cho *et al*, 2015). Multiplexed immunological methods were recently developed but are associated to more expensive equipment (Black *et al*, 2019).

In conclusion, immunological methods and more particularly ELISA methods, are the most widely used approach to detect and quantify allergens because they are sensitive and specific for the detection of allergenic proteins, cost effective and easy to use. However, cross-reactivity with other proteins that display high homology sequences and false negative results induced by food processing remain the major limitations of these methods.

### 2.4. Mass spectrometry-based methods

As we have seen, immunological and DNA-based methods for the detection and quantification of food allergens present a series of limitations. During the last decade, alternative methods have been developed with a particular emphasis placed on various mass spectrometry approaches. Coupled to liquid chromatographic separation, mass spectrometry was successfully used for food allergen **identification/characterization**, and more recently in allergen **quantification** as well (Monaci *et al*, 2018; Planque *et al*, 2019).

Mass spectrometry can be used for two main tasks related to food allergen analysis: characterization/identification and detection/quantification (Figure 8). These two tasks are closely related and are both required for the development of quantitative analysis methods.

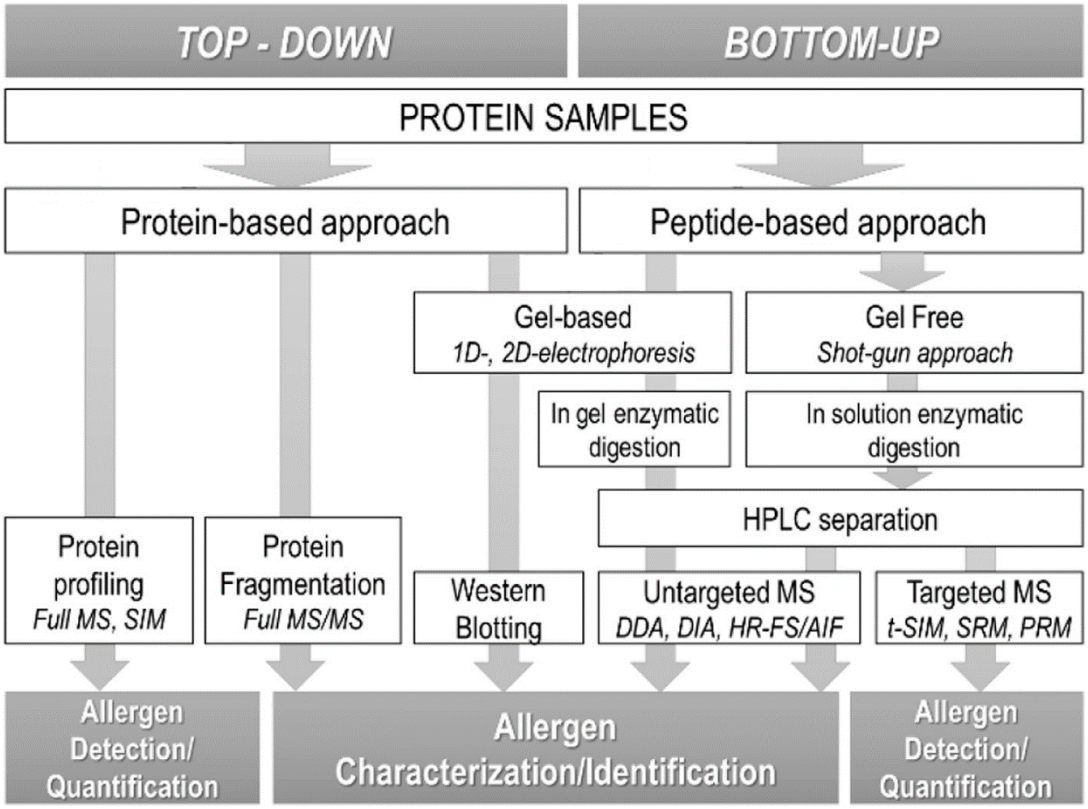



*Figure 8 - Summary of the different MS-based analytical approaches available for food allergen identification, characterization, and quantitative detection* (adapted from Monaci *et al*, 2018).
*Abbreviations: MS = mass spectrometry; SIM = selected ion monitoring; MS/MS = tandem mass spectrometry; HPLC = high-performance liquid chromatography; DDA = data dependent analysis; DIA = data independent analysis; HR-FS/AIF = high resolution full scan/all-ion fragmentation.*

Food allergen characterization and identification are essential preliminary steps in the development of analysis methods. Food ingredient proteome, protein sequences and structures, different proteoforms and isoforms for a given protein, post-translational modifications or parameters of food processing-induced modifications are examples of information that can be gathered (Andjelković *et al*, 2017). These aspects are essential for biomarker selection, the targets of the analysis method. This question of biomarker selection is extensively described in a dedicated section in this document (see page 43, "Identification of peptide biomarkers for the detection and quantification by UHPLC-MS/MS of four allergenic ingredients (egg, milk, peanut and hazelnut) in processed food products").

The detection and quantification of food allergens can be achieved with a protein-based (top-down) or a peptide-based (bottom-up) approach.

Food allergen quantification with a **protein-based approach** may be obtained by directly spraying the protein solution using electrospray. Potentially, a protein solid phase extraction and separation by liquid chromatography can also be implemented before as in the work of Monaci and Van Hengel (Monaci & Van Hengel, 2008). These authors developed a method to quantify cow's milk proteins in mixed fruit juice samples with estimated MDL and MQL of 1 and 4 µg/ml respectively. Immunocapture with magnetic particles covered with specific antibody is another approach to purify proteins before mass spectrometry analysis. Schneider and co-workers developed a method to detect lysozyme, a hen egg with protein, in cheese at a level of 5 mg/kg (Schneider *et al*, 2010).

Despite the good results obtained in these studies (in terms of sensitivity in the order of the ppm and recovery), the quantification of intact proteins in complex matrices such as processed food is extremely

complicated and suffers from several limitations. Indeed, ion suppression (competition for ionisation efficiency in the ionisation source) or high complexity and superposition of numerous peaks in the mass spectra in the presence of different proteins are limiting the sensitivity of top-down methods (Monaci *et al*, 2018). Moreover, the related spectra are highly complex and characterized by a wide distribution of the multiple charge states of the protein of interest. This characteristic multi-charge cluster can be influenced by food processing (such as thermal treatment or the presence of other matrix compounds), affecting therefore accuracy of quantitative methods (Pilolli *et al*, 2020). Finally, the protein-based approach is also limited by the commercial availability of specific purified protein standards and internal standards. For these reasons, mass spectrometry-based methods for the detection and quantification of food allergens are using a bottom-up approach, targeting specific peptide biomarkers obtained from protein enzymatic digestion (Monaci *et al*, 2018).

As presented in Figure 9, the development of a mass spectrometry-based method using the **bottom-up** approach for the quantification of food allergens can be divided in a succession of different steps.

The first step in the development of a mass spectrometry-based method is the identification and further selection of possible peptides and thus, proteins, specific to the studied species (i.e., biomarkers). This selection is particularly important when developing a quantitative method since quantifying the allergenic ingredient will be directly linked to quantifying the selected biomarkers. Therefore, the overall accuracy, robustness, and precision of the method rely on this selection. Two potential strategies can be implemented for **peptide biomarkers selection**.

The first strategy is based on an *in silico* peptide selection (Figure 9 Route A) combining the search in protein databases such as UniProt (http://www.uniprot.org/) using a software such as Skyline (MacLean *et al*, 2010) to perform *in silico* peptide digestion and work out a mass spectrometry method (Planque *et al*, 2017a). The main advantage of this strategy is that it can be performed on **triple quadrupole mass spectrometers** and does not require a high-resolution mass spectrometer which is more expensive and rather designed for research instead of routine analyses. However, this strategy is dependent on the availability of the information in protein databases and is time consuming when numerous peptides have to be screened. However, this strategy is dependent on the availability of the information in protein databases and is time consuming when numerous peptides have to be screened.

The second strategy is based on an instrumental approach with **high resolution mass spectrometry** (HRMS). The sample is analysed with an untargeted approach (Figure 9 Route B). Acquired data are processed with an algorithm (MASCOT, X!Tandem, SEQUEST) to be assigned to peptides obtained from a protein database. This strategy is also dependent on the availability of information in protein databases but has the advantage to characterize the global peptide and protein profiles for a given sample in a single analysis.

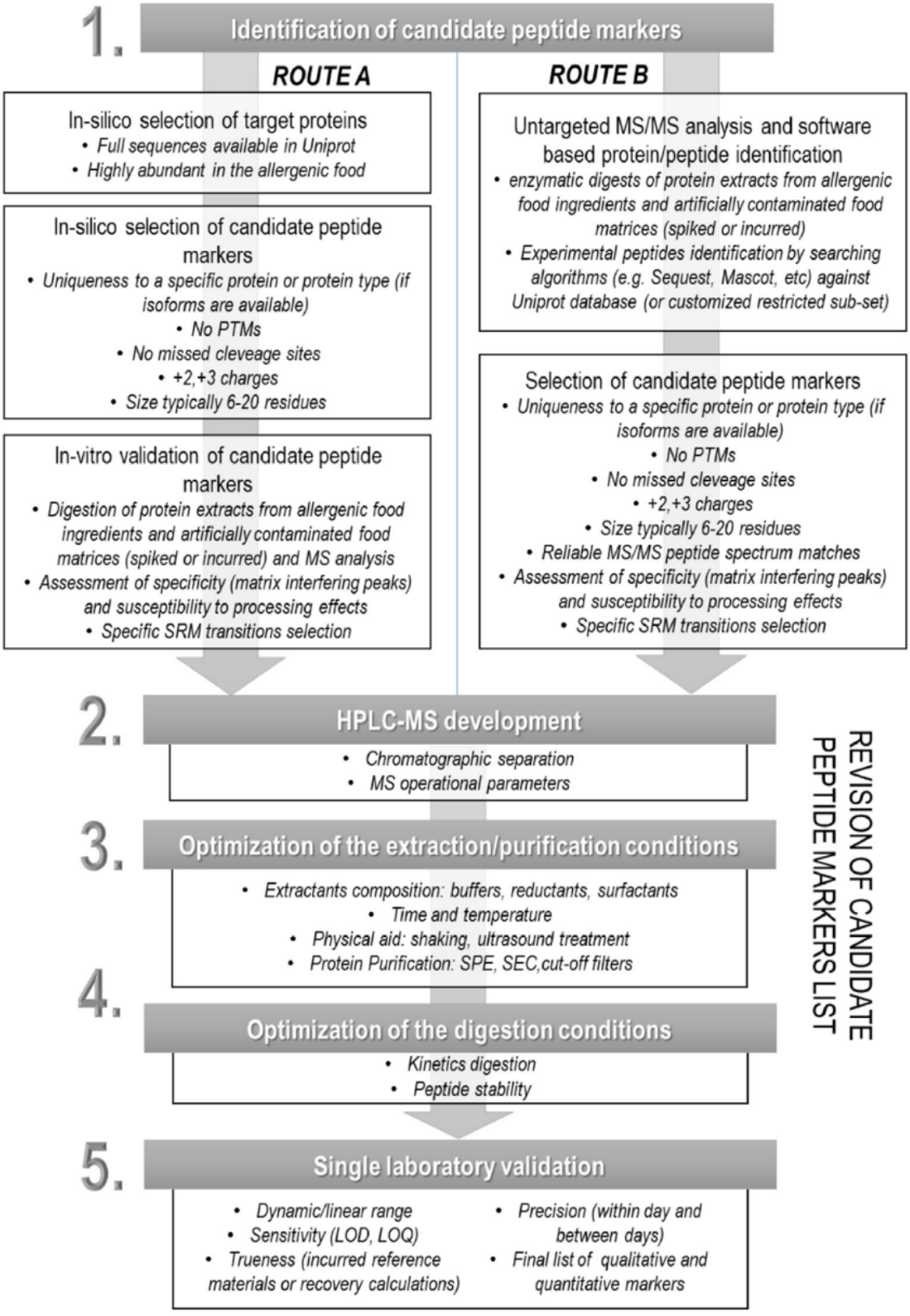



*Figure 9 - Workflow in the development of mass spectrometry based-method for the detection of food allergens using a bottom-up strategy* (Monaci *et al*, 2018).

These two strategies allow the identification of potential peptide biomarkers but the selection of ideal peptide biomarkers is based on several criteria to guaranty the performance of the quantitative method. Criteria such as resistance to food processing, specificity of the peptides for a considered food

ingredient, effects of amino acid modifications, peptide length or enzymatic digestion efficiency also need to be considered. The application of these criteria for the selection of peptide biomarkers for egg, milk, peanut and hazelnut is extensively described in a dedicated section on page 43 ("Identification of peptide biomarkers for the detection and quantification by UHPLC-MS/MS of four allergenic ingredients (egg, milk, peanut and hazelnut) in processed food products").

The second step in the development of mass spectrometry-based methods for allergen quantification is the instrumental **liquid chromatographic separation** and mass spectrometry analysis for the previously identified peptides biomarkers. Given the high complexity of the various and different samples, peptides need first to be separated before mass spectrometry analysis, to decrease competition between peptides during ionization and to reduce the likelihood of ion suppression, increasing the possibility for the detection of low-abundance peptides (Di Palma *et al*, 2012). The common approach is based on reversed-phase chromatography with a separation of peptides based on their hydrophobicity.

Although several quantitative proteomic techniques have been reported (Monaci *et al*, 2018), the most extensively used targeted technique is **selected reaction monitoring (SRM)**, also known as multiple reaction monitoring (MRM) when applied to multiple product ions from one or more precursor ions (Croote & Quake, 2016). In this technique, tens to hundreds of peptides can be quantified with a high sensitivity (ppm level, comparable to ELISA). The triple quadrupole mass spectrometer underlies SRM, where the first and last quadrupoles act as static mass filters for a precursor and product ions, respectively, and the second quadrupole functions as a collision cell to fragment the precursor ion into product ions (Figure 10).

Each target in an SRM assay is known as a transition and consists in the mass-to-charge ratio (m/z) of a precursor ion and one of its product ions. The combination of chromatographic retention time, quadrupole resolution (ability to distinguish two ion species, typically m/z = 1 for quadrupoles) and the use of multiple transitions confer a high specificity to these analytical methods (Kushnir *et al*, 2005).

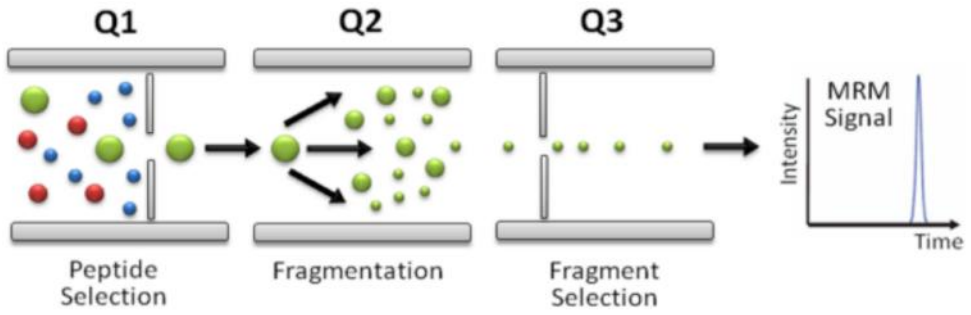



*Figure 10 – Schematic visualization of SRM analysis of peptides in a triple quadrupole mass spectrometer* (Schmidt *et al*, 2008)

The optimization of the different steps of the **sample preparation** is another critical feature in the development of detection and quantitative methods. These steps include at least **protein extraction** and their **enzymatic digestion**. Trypsin is by far the preferred enzyme in proteomics due to its high catalytic activity, its specificity and the ideal average mass range for mass spectrometry analysis and good ionization properties of generated peptides (Bunkenborg *et al*, 2013). Denaturation of the proteins with the reduction and alkylation of disulfide bridges is often performed before the digestion step to increase enzyme accessibility to the digestion sites. Finally, **purification steps** at the protein and/or peptides levels are most of the time required for successful analysis and to eliminate interfering components (from sample preparation buffers and food matrix) (Planque *et al*, 2017a).

The approach used for the sample preparation directly affects the detection of the selected peptide biomarkers. In the case of multiple food allergen detections, it is inconceivable that a single sample preparation protocol (extraction, digestion and purification) could be optimized for the detection of multiple peptides from multiple proteins, characterized by specific and different extractability and digestibility (Johnson *et al*, 2011). For multiple allergens analysis, a compromise on method performance (sensitivity and accuracy) for the different peptides must necessarily be considered. A compromise and potential strategy is thus to identify and preselect numerous potential peptide biomarkers. Sample preparation optimization is evaluated based on this preselection and only the best performer peptide biomarkers are kept in the final analysis method.

The last step in the development of a method for food allergen analysis is the **validation**. This step is required to assure that the method is suitable for its intended purpose and that the same method will perform equally in all laboratories trying to repeat the analysis. For that purpose, the Association of Official Analytical Chemists, **AOAC International**, published a document in 2016 entitled "**Standard Method Performance Requirements (SMPRs®)** for Detection and Quantitation of Selected Food Allergens" (Paez *et al*, 2016). This document describes the minimum recommended performance characteristics to be achieved for the detection and quantification by mass spectrometry of egg, milk, peanut, and hazelnut food allergens in finished food products and ingredients. Required values for analytical parameters including method quantification limit (MQL), method detection limit (MDL), repeatability with a repeatability relative standard deviation ($RSD_r$), reproducibility with a reproducibility relative standard deviation ($RSD_R$) and recovery are defined in the document (see Table 3).

However, the complete validation of a method for multiple allergen quantification is currently hard to achieve in the absence of proper **reference materials**. Specific allergenic ingredients and incurred matrices, supplied with characterization of total protein content, allergen profiling, homogeneity and stability tests are needed for coherent method validation taking the impact of food processing into account. Indeed, **food processing** and **food matrix** are known to modify food allergen detection as they affect, for example, the protein extractability and/or digestibility (Korte *et al*, 2019). The different strategies to compensate for these matrix effects and develop absolute allergen quantification methods are discussed in the next section, especially the use of **stable isotope-labelled internal standards**.

*Table 3 – Standard method performance requirements for detection and quantification of selected food allergens by mass spectrometry (AOAC SMPR 2016.002)*

| | **Target allergen** | | | |
|---|---|---|---|---|
| **Parameter** | Whole egg | Milk | Peanut | Hazelnut |
| Analytical range (ppm) | 10-1000 | 10-1000 | 10-1000 | 10-1000 |
| MQL (ppm) | ≤ 5 | ≤ 10 | ≤ 10 | ≤ 10 |
| MDL (ppm) | ≤ 1.65 | ≤ 3 | ≤ 3 | ≤ 3 |
| Recovery (%) | 60 - 120 | 60 - 120 | 60 - 120 | 60 - 120 |
| $RSD_r$ (%) | ≤ 20 | ≤ 20 | ≤ 20 | ≤ 20 |
| $RSD_R$ (%) | ≤ 30 | ≤ 30 | ≤ 30 | ≤ 30 |

In conclusion, mass spectrometry-based methods are very powerful techniques for food allergen analysis with high accuracy, sensitivity, specificity and reproducibility. Additionally, this technology is not dependent on biological interactions between an antibody and the allergen potentially affected by food processing-induced protein denaturation and/or cross-reactivity with other protein issues. The current limitations of mass spectrometry-based methods are the need of highly trained technicians and expensive equipment. Analysis delay is also longer compared to immunological and DNA-based methods, but potentially counterbalanced by the multiplex analysis.

Two main strategies are possible for food **allergen quantification** by mass spectrometry, label-free or isotopic dilution.

The **label-free** quantification strategy is based on the comparison of peptide signal intensity between different samples by using a calibration curve (Mattarozzi *et al*, 2012). This approach requires a calibration curve for each matrix to be analysed and is particularly well adapted when a limited number of matrices are encountered, which is not the case in food allergen analysis. Other label-free quantification alternatives, based on modified synthetic peptides (Zhang *et al*, 2012) or on standard addition (Posada-Ayala *et al*, 2015), are found in the literature but as anecdotal evidence.

**Isotope dilution** mass spectrometry is, most likely, one of the best analytical techniques to provide reliable MS-derived absolute quantitative data (Villanueva *et al*, 2014). The isotopically labelled analogue to the targeted compounds, namely the **internal standard**, is added to the sample and simultaneously analysed with the analyte of interest. This internal standard is used to correct for variability generated by the analysis procedure itself, from sample preparation to LC-MS/MS analysis. The internal standard is also used to minimise changes in the analytical performance over time and between laboratories.

Food matrix and production history are also known to affect peptide detectability (Korte *et al*, 2019) as they affect, for example, the protein extractability and might interfere with enzymatic digestion. The amino acid protein sequence is also known to influence kinetics of tryptic digestion and some of them (such as two consecutive arginine or lysine residues) can lead to low rate or incomplete digestion (Villanueva *et al*, 2014). Another example of the impact of food matrix is the ionization suppression effect resulting from competition for ionisation efficiency in the ionisation source between the analyte and co-eluting compounds (Rodríguez-González & Ignacio García Alonso, 2019). All these **matrix effects** could lead to biased allergen quantifications and isotopically labelled internal standard is used to correct for variability, depending on the selected form, some of these effects.

In food allergen analysis, the initial analytes are the proteins of the allergenic ingredient. However, with the SRM approach, food allergens are detected through signature peptides. Different forms of isotope labelled internal standards are possible and proposed in the literature (Planque *et al*, 2017a).

The **gold standard** for the absolute food allergen quantification is the use of the isotopically labelled version of the analysed protein. In this approach, the internal standard accounts for variations during the whole sample preparation and LC-MS/MS analysis. This technique was used by Newsome and Scholl to quantify allergenic bovine milk $\alpha_{S1}$-casein in baked goods (Newsome & Scholl, 2013). The use of such internal standards increases the accuracy of the allergen concentration measured in baked goods by correcting for extraction recovery. However, these authors found that the SRM assay underestimated concentrations because the allergen protein was transformed by food processing that altered peptide detection. This approach is limited by the fact that **labelled proteins** are very often not commercially available and custom synthesis is financially unaffordable for routine laboratories (Planque *et al*, 2017a).

An alternative and relatively **affordable** approach described in most of the studies relies on isotopically labelled **synthetic peptides**. However, synthetic peptides are not part of all the steps of the sample preparation, as they escape to the enzymatic digestion and are, therefore, not able to correct for the variability introduced at this step of the procedure (Planque *et al*, 2017b).

The use of "winged peptides" or "long peptides" was proposed to circumvent the "digestion" limitations of isotopically labelled synthetic peptides. In this strategy, the internal standard is composed of the isotopically labelled analogue of the peptide biomarker flanked at each end by a couple of amino acid residues of the corresponding natural protein sequence. With this approach, the internal standard, as the analyte, needs to be digested. This strategy, correcting for digestion-step-related effects, improves allergen quantification when compared to the use of classical synthetic peptides (Chen *et al*, 2016; Zhang *et al*, 2014b).

Another strategy, known for more than a decade in the proteomic field but never applied to food allergen analysis, is the use of **concatemer** (Pratt *et al*, 2006) also known as QconCAT. This strategy is based on the use of an isotopically labelled artificial/chimeric protein, recombinantly produced, and composed of concatenated peptide biomarkers. In this approach, the variability of the enzymatic digestion is thus taken in to account since this type of internal standards needs to be digested to release its constituting peptides (Brownridge *et al*, 2012). The advantage of this method is that a single concatemer can contain peptides belonging to different proteins or allergens and can therefore be used as a single internal standard for multiple food allergen analysis. With this multiplexed property, the use of concatemer is cost-effective when compared to synthetic peptides. This approach was implemented in this thesis and is extensively described in a dedicated chapter (see page 127:"Development, production, characterization and evaluation of the performance of the analytical method developed using stable isotope-labelled internal standards ($^{15}$N concatemer and $^{15}$N β-lactoglobulin)").

## 2.5. Conclusion

The statements of Holzhauser and Röder in 2015 and Walker and co-workers in 2016 mentioned hereunder, remain timely, providing appropriate perspectives for food allergen analysis and they highlight the main challenges that analytical chemists will face in the future.

*"Market requirements and operator demands will largely define the future of allergen detection. One conclusion is, however, clear: There will not be one single type of method that combines all economic and analytical demands – an easy-to-use, rapid, sensitive, specific, quantitative, on-site, multi-target and cheap allergen detection for any allergenic food component in any analytical environment"* (Holzhauser & Röder, 2015).

*"Legislation, risk assessment and risk management of food allergens show a high dependency on the ability to detect food allergens and quantitatively determine them. All current analytical approaches exhibit described deficiencies that jeopardise accurate results and risk false positives and false negatives. If we fail to realise the promise of many strands of risk assessment and risk management of food allergens through lack of the ability to measure food allergens reproducibly and with traceability, the analytical community will have failed a significant societal challenge"* (Walker *et al*, 2016).

It is in this context that the PhD thesis undertaken during the "Allersens" project was designed and implemented. More specifically, the aim of this project was to develop and validate a multi-allergen mass spectrometry-based method targeting 4 priority allergens (peanut, hazelnut, egg and milk) and

being robust to food processing. We will now rapidly describe the different tasks and work packages of this research project in which the PhD thesis work was done.

## 3. The "Allersens" project

This thesis was part of the "Allersens" project, a thematic call project funded by the Federal Public Service Health, Food Chain Safety and Environment (FPS Health). This project involves different partners, the University of Namur and two research centres, CER Groupe and ILVO. We were two PhD students in charge of the implementation of this project with complementary tasks. The present manuscript is focused on development and results obtained by myself at the University of Namur and at CER Groupe. However, for reasons of consistency, results obtained by both PhD students will be presented. Results obtained at ILVO by Kaatje Van Vlierberghe, the other PhD student, will always be clearly referenced.

The "Allersens" project is entitled "*Development and validation of a quantitative LC-MS/MS method as a reference for the detection of multiple allergens (hazelnuts, peanuts, milk, and eggs) in processed food products – Comparison with existing methods (ELISA, PCR,…) and new technologies (flow cytometry, droplet PCR, …)*". The project is divided in five interconnected work packages (WP), corresponding to well defined tasks (Figure 11).

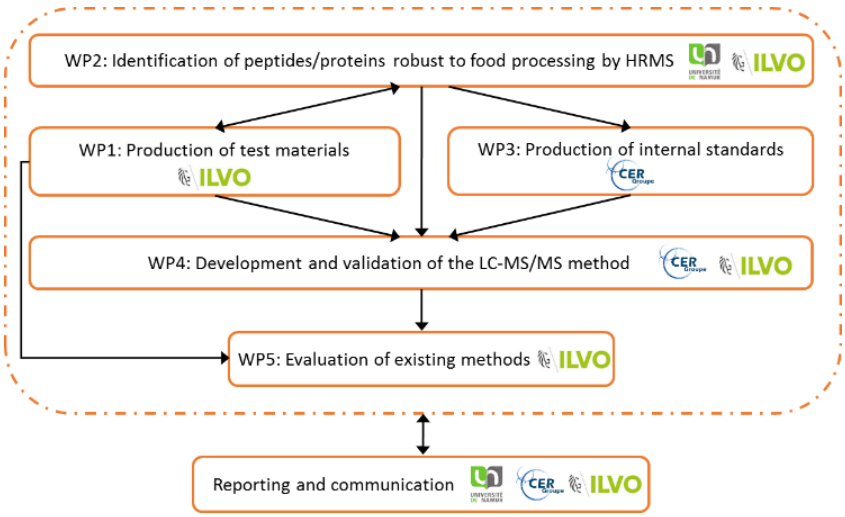



*Figure 11 - Schematic overview of the "Allersens" project and work packages (WP) distributions between the different institutions (University of Namur, CER Groupe and ILVO)*

### 3.1. Production of test material (WP1)

It is known that different food processing conditions and matrix interactions have effects on the detectability and/or recovery of target molecules during the analysis (Korte *et al*, 2019). To include these aspects in the development of the method, different food matrices, containing hazelnut, peanut, milk or egg, were produced under well-defined and controlled conditions. The food matrices and processing conditions were selected in function of representativeness for real food products and in function of processing conditions that are applied nowadays in the food industry. The matrices and test materials were produced during the "Allersens" project by Kaatje Van Vlierberghe in the Food Pilot unit of ILVO, a semi-industrial environment, with well-defined and monitored conditions. Produced test materials were used for several tasks in the project, including selection of peptide biomarkers, development and validation of the quantitative UHPLC-MS/MS method and for the comparison of the developed method with existing methods.

### 3.2. Identification of peptides/proteins robust to food processing by HRMS (WP2)

The first step in the development of a MS-based method is the identification of biomarkers, which are, in the case of food allergens, peptides. The objective of this work package was to identify, through an empirical approach based on HRMS, potential peptide biomarkers for the detection and quantification of food allergens in processed food products. Food processing and matrix interactions were considered with the analysis of test materials, separately containing one allergenic ingredient and submitted to different food processing techniques.

The four allergenic ingredients considered in this project were separated between the two PhD students. Kaatje Van Vlierberghe was in charge of the identification of potential peptide biomarkers for milk and hazelnut whereas I worked on egg and peanut.

### 3.3. Production of internal standards (WP3)

Mass spectrometry quantification is mostly performed with the stable isotope dilution method. In this method, a stable isotope-labelled internal standard is spiked in the sample to compensate for the loss of analyte and matrix effects during sample preparation and MS analysis (Villanueva *et al*, 2014). The ideal internal standard for mass spectrometry is a stable isotope-labelled analogue of the quantified analyte.

The common approach is based on the use of isotopically labelled synthetic peptides. Although easily available from commercial providers, relatively cheap and easy to use, synthetic peptides suffer from several limitations since the initial analyte is the protein from the allergenic ingredient. Therefore, synthetic peptides do not compensate for protein loss during extraction and clean-up or for uncomplete enzymatic digestion. Another, and theoretically, ideal strategy is based on the use of isotopically labelled proteins, compensating for the different steps of the sample preparation and MS analysis. However, beside technical limitations, this solution is currently inconceivable for routine laboratories due to the cost of this technology and the need for at least one isotopically labelled protein per analysed allergenic ingredient.

The strategy considered in this thesis is based on the use of stable isotope-labelled concatemers. A concatemer is an artificial/chimeric protein, recombinantly produced in an environment that allows labelling with stable isotopes, and composed of several concatenated peptide biomarkers for the four targeted allergenic ingredients. This technology is known for more than a decade in the proteomic area but has never been applied to food allergen analysis. In contrast to synthetic peptides, concatemers need to be proteolytically digested to release their peptides, and thus, this peptide release is also affected by the interference caused by the matrix during the digestion step, in a manner similar to the analyte of interest. Another advantage of concatemers is their potential for multiplexing. A single concatemer can be composed of numerous proteotypic peptides and can therefore be used for multiplexed allergen analysis and be cost-effective when compared to synthetic peptides.

Based on peptide biomarkers selected in WP2, the concatemer design and production was entirely performed at CER Groupe by myself.

### 3.4. Development and validation of the LC-MS/MS method (WP4)

The goal of this work package was to develop and validate a quantitative UHPLC-MS/MS method for the 4 targeted allergenic ingredients (egg, milk, peanut and hazelnut). The development of the method included optimization of both sample preparation protocol (extraction/purification/digestion) and

UHPLC-MS/MS parameters. This work package gathered outcomes of the three first ones. The UHPLC-MS/MS method was developed and optimized with the samples produced in WP1, with the targeted peptide biomarkers selected in WP2 and was worked out as quantitative thanks to the use of the isotopically labelled internal standard produced in WP3.

Sample preparation optimization was already addressed for peptide biomarkers identification in WP2 and was complementary accomplished by both Kaatje Van Vlierberghe and myself. Performance of the developed method was further evaluated through a complete validation in both, ILVO and CER Groupe laboratories. This validation was based on AOAC guidelines for food allergen quantification by mass spectrometry (Paez *et al*, 2016).

### 3.5. Evaluation of existing methods (WP5)

The goal of this last work package is to compare the developed LC-MS/MS method with other food allergens analysis methods including ELISA, PCR and flow cytometry. This task will be performed by Kaatje Van Vlierberghe, in the second half of 2020 and will not be part of this thesis.

## 4. Objectives of the thesis

As described in the previous section, this thesis was included in the "Allersens" project, aiming to develop and validate a quantitative LC-MS/MS method for the detection of multiple allergens (hazelnuts, peanuts, milk, and eggs) in processed food products. Three main tasks of this project were addressed in this thesis.

The first one was the identification of potential peptide biomarkers for the detection and quantification of egg and peanut by mass spectrometry in processed food products. The identification of these potential peptide biomarkers was performed using an empirical approach based on HRMS and the analysis of test materials, separately containing one allergenic ingredient, and prepared by Kaatje Van Vlierberghe at ILVO. Different food processing techniques (cooking and induction of Maillard reactions) and matrix composition and properties (fatty rich, low pH or complex as chocolate with the presence of tannins and polyphenols) were investigated to consider their impact on allergen detectability.

The second addressed task was the development and the production, as a recombinant protein, of a stable isotope labelled concatemer used as internal standard in the quantitative UHPLC-MS/MS method. The proof of concept and several optimisations were performed with concatemers composed of egg peptides. The pursued objectives are a sufficient expression yield to be cost effective compared to traditionally used synthetic peptides and a sufficient isotopic enrichment to avoid the introduction of false positive results with the use of the concatemer in routine analysis. The optimized protocol was finally applied to produce a concatemer containing peptide biomarkers from the four considered allergenic ingredients (egg, milk, peanut and hazelnut).

The final task of the thesis was the complete validation of the quantitative UHPLC-MS/MS method targeting the four allergenic ingredients. The method was developed by Kaatje Van Vlierberghe and based on the HRMS selected peptide biomarkers. The method was validated combining the use of the $^{15}N$ stable isotope-labelled concatemer as internal standard and standard addition method for quantification. The validation was based on AOAC guidelines for food allergen quantification by mass spectrometry (Paez *et al*, 2016).

**RESULTS**

**PART I**

**Identification of peptide biomarkers for the detection and quantification by UHPLC-MS/MS of four allergenic ingredients (egg, milk, peanut and hazelnut) in processed food products**

## 1. Objectives

The overall aim of this project was to develop and validate a quantitative UHPLC-MS/MS method for the detection of multiple allergenic ingredients (hazelnuts, peanuts, milk, and eggs) in processed food products. As already mentioned, the first step in the development of any mass spectrometry-based method is to identify the targets of the analysis. In the case of food allergen detection by MS, the analytes are constitutive peptides derived from proteins present in these food ingredients. The objectives of this first task was thus to identify peptides that can be potential biomarkers for the quantitative analysis of the four allergenic ingredients. This selection is particularly important when developing a quantitative method as the quantification of the allergenic ingredients will be directly and positively correlated to the quantification of the selected biomarkers. Therefore, the method overall performance, including specificity, accuracy and robustness, relies on this selection.

To identify potential peptide biomarkers, a bottom-up empirical approach using HRMS was chosen. Such a bottom-up proteomic strategy consists usually in extracting the proteins from the matrix followed by a protein digestion using enzymes such as trypsin. The resulting peptides are then separated by liquid chromatography and analysed by mass spectrometry. With the data obtained, Mascot search engine is used in order to identify the detected peptides (Figure 12). To increase the robustness of the peptide biomarker selection and take into account food processing, different food matrices, separately containing hazelnut, peanut, milk or egg, were produced under well-defined and controlled conditions.

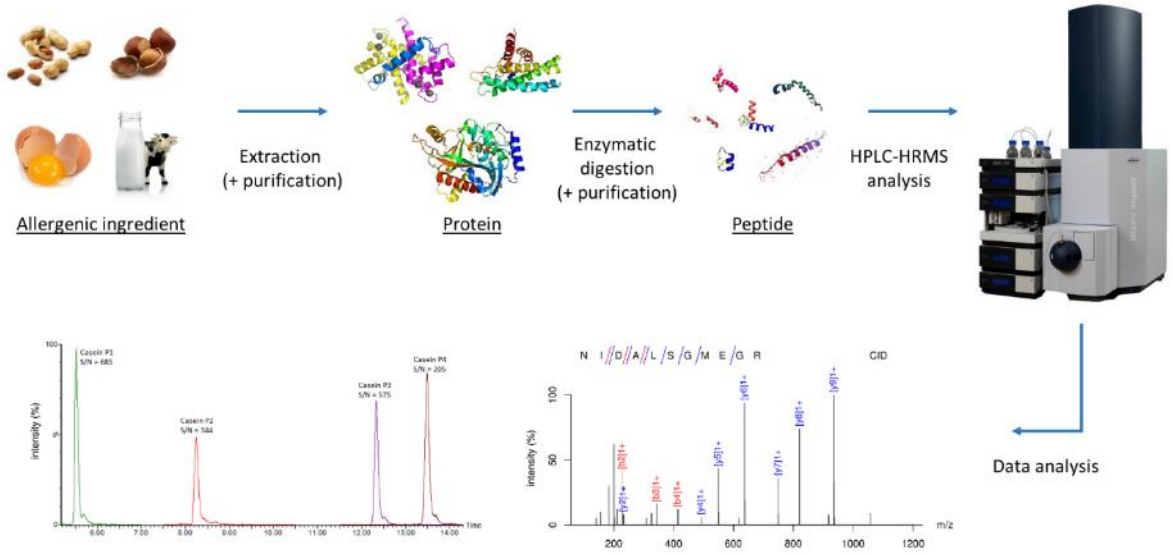



*Figure 12 - Bottom-up proteomic approach used to identify peptide biomarkers. Proteins of the allergenic ingredient are extracted from the matrix and enzymatically digested into peptides. The resulting peptides are analysed by HPLC-HRMS and identified with the use of Mascot search engine.*

Sample preparation is the starting point of the analysis. There is no one standard method for sample preparation. Protocols differ according to the sample types and experimental goals such as peptide identification. An efficient identification is only possible if proteins are correctly extracted and enzymatically digested. The optimization of a protocol was therefore a preliminary step in the selection of peptide biomarkers. Starting from literature analysis and previous experience of the different

partners, especially from the CER/UNamur “Allermass” project (M. Planque et al., 2016), a sample preparation protocol was developed and optimized for HRMS analysis. This protocol was finally systematically applied to all food matrices to identify potential peptide biomarkers for the quantitative analysis of egg, milk, peanut and hazelnut in processed food products.

## 2. Production of test matrices (entirely done by Kaatje Van Vlierberghe at ILVO)

As mentioned, different food processing conditions and matrix interactions may have an impact on the detectability and/or recovery of target molecules during the analysis. To include this variability in the selection of peptide biomarkers and identify the robust ones, several test matrices were produced. These matrices separately contained each one of the four allergenic ingredients (egg, milk, peanut and hazelnut) and were subjected to four different processing techniques. In addition, the unprocessed form of these allergenic ingredients was also analysed. Processing conditions that were applied on the allergenic ingredients are: roasting/heating, induction of Maillard reactions (chemical reactions between amino acids and reducing sugars at high temperature), mixing in a fatty rich matrix or in a matrix with low pH. The food matrices and processing conditions were selected according to their representativeness for “real” food products and in function of processing conditions that are currently applied in the food industry (Fellows, 2009) (see Figure 13 for egg, Figure 14 for milk and Figure 15 for peanut and hazelnut).

To include potential variability in protein content and protein distribution due to growing conditions and origin, peanut and hazelnut from two distinct geographical growing locations were considered for the production of test materials. Hazelnuts from Italy and Turkey were selected. Peanuts corresponding to two market types, the common peanut classification, were selected and were originating from China and Israel. Consequently, each test matrix was produced in duplicate, one for each peanut and hazelnut origin.

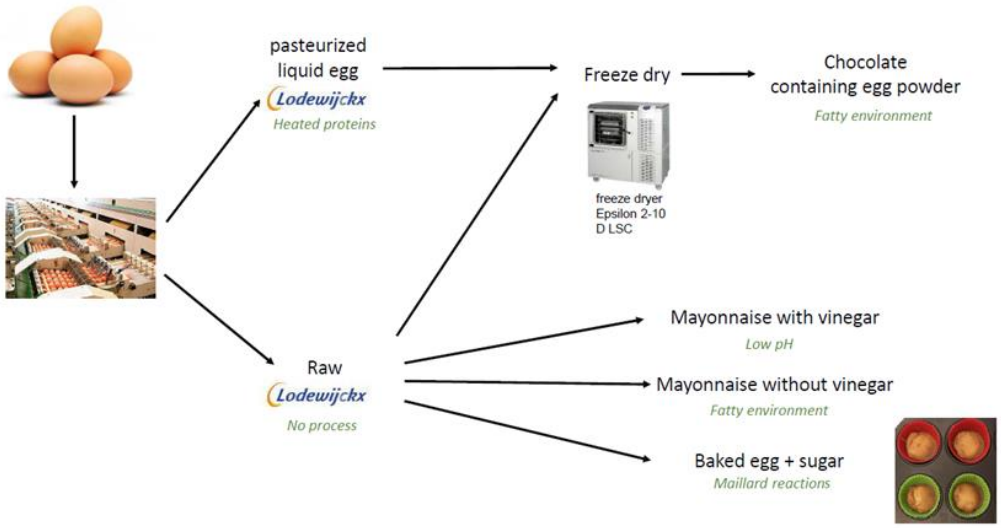



*Figure 13 - Schematic overview of processing conditions applied to egg to include the associated variability in the identification of peptide biomarkers*

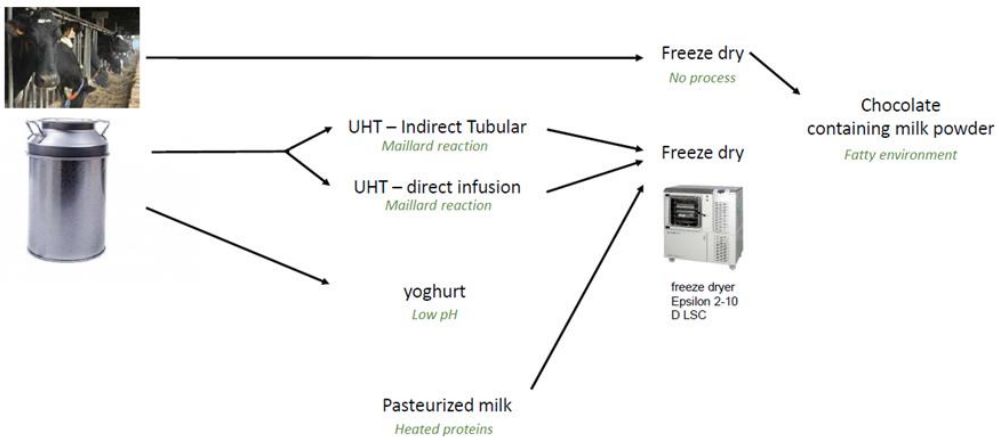



*Figure 14 - Schematic overview of processing conditions applied to milk to include the associated variability in the identification of peptide biomarkers*

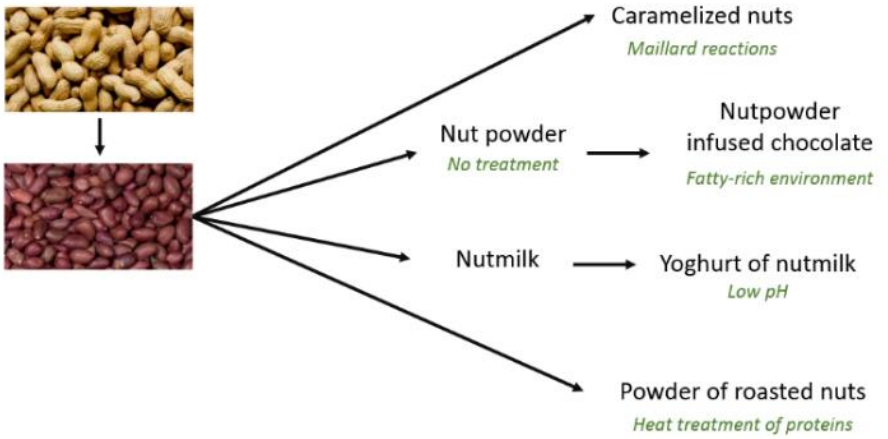



*Figure 15 - Schematic overview of processing conditions applied to peanut and hazelnut to include the associated variability in the identification of peptide biomarkers. Each matrix was produced in duplicate since peanut and hazelnut from two geographical origins were considered to evaluate potential variation at the protein level.*

All processing was carried out by Kaatje Van Vlierberghe in the Food Pilot with amounts that allowed producing homogenous testing samples which were representative for real live products. During processing, all relevant processes (temperature, cooking time, pH) and product quality parameters (origin, purity) were constantly monitored. This detailed monitoring of the production process allowed us to guarantee the composition of the reference material and enables, when needed, the production of new batches of the reference material under exactly the same conditions to assure homogeneity/quality and standardization of the reference material production. The resulting products were packaged under vacuum and stored in the dark at -80 °C until use.

## 3. Sample preparation optimization

Food allergen bottom-up proteomic approach is based on the analysis of specific peptides obtained from enzymatic digestion of proteins of the allergenic ingredient of interest. The two main steps in the sample preparation are the extraction of proteins from the matrix and the enzymatic digestion of these proteins to generate peptides. Purification steps, at the protein and the peptide levels, can also be necessary. Indeed, the enzymatic digestion requires specific conditions for optimal enzyme activity (respect of optimal enzyme pH, low level of denaturing agents,…), and some components can interfere or are not compatible with mass spectrometry analysis (the presence of non-volatile salts, some detergents such as Triton X-100 or NP-40). All these criteria were evaluated and optimized based on measurable parameters such as the protein extraction yield, the number of identified peptides and proteins by HRMS and/or the signal intensities associated with the identified peptides. Practical criterions such as duration and cost of the sample preparation protocol were also considered to ensure its transferability to a routine method.

### 3.1. Optimization of the protein extraction protocol (by Kaatje Van Vlierberghe at ILVO)

Optimization of protein extraction and of sample purification at the protein level was performed by Kaatje Van Vlierberghe at ILVO based on the protocol developed during the "Allermass" project (Planque et al., 2016). Here is the summary of obtained results:

- The extraction buffer developed in the "Allermass" project (2 M Urea in 50 mM Tris-HCl; pH 9.2) presented optimal extraction yields when compared to extraction buffers employed in partner laboratories and food allergen extraction buffers identified in the literature (Figure 16).
- The protein precipitation by methanol/chloroform used to purify the proteins resulted in a significant protein loss and was not retained for future analysis (Figure 16).

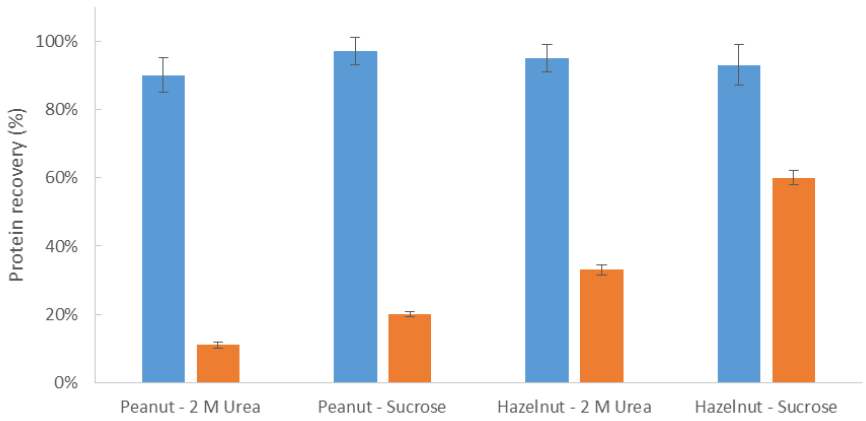
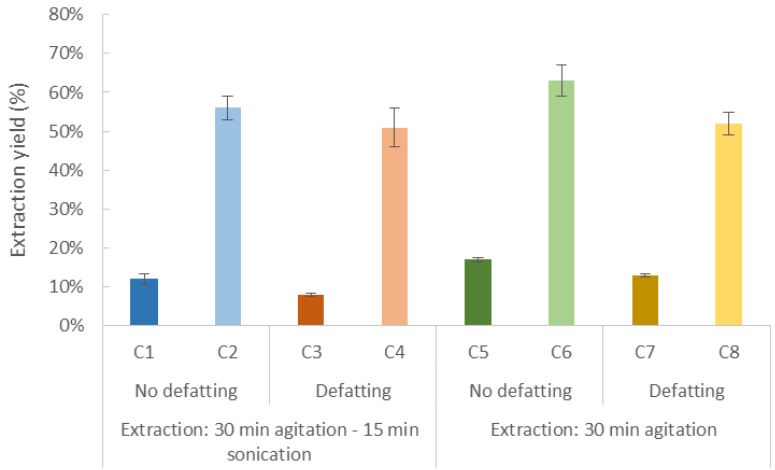


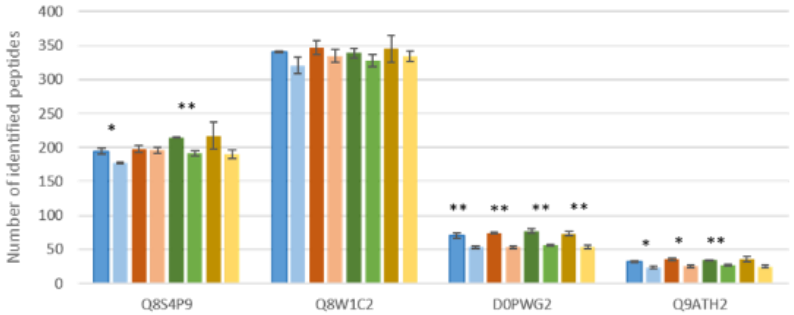

*Figure 16 - Example of comparison of protein recovery, in raw peanut and hazelnut sample, using different extraction buffers ("Allermass" project buffer (2 M Urea in 50 mM Tris-HCl; pH 9.2) and "sucrose" buffer form ILVO (30 % sucrose, 0,1 M KCl, 50 mM Tris-HCl, 5 mM EDTA, 1 mM DTT; pH 8)), with (in orange) and without (in blue) methanol/chloroform protein precipitation. Results are expressed as protein recovery and represent the means ± 1 S.D. for 3 independent experiments. Proteins were measured by monitoring the absorbance at 280 nm and protein recovery was calculated based on the theoretical protein content of peanut (25 %) and hazelnut (15 %).*

- The extraction procedure developed in the “Allermass” project, combining 30 min of agitation and 15 min sonication, showed optimal performance when compared to longer extraction procedures (Figure 17).
- A sample defatting step (with three repetitions of 20 ml of n-hexane addition followed by centrifugation and removal of the supernatant) prior to the protein extraction is a time and reagent consuming step. No significant effect on protein and peptide identification (Figure 18) or on peptide MS signal intensity was observed (Figure 19). This step was therefore not retained in our future analyses.
- The urea concentration needs to be reduced to ensure efficient tryptic digestion. This can be achieved by sample desalting or dilution. Sample desalting by gel filtration resulted in a significant quantitative loss of protein (Figure 17). However, qualitatively, the same major proteins and peptides were identified when compared to the diluted extract (Figure 18). When performing trypsin digestion on the same amount of protein in both protocols, the desalting resulted in higher intensities of peaks for the identified peptides (Figure 19). The sample desalting was therefore considered and included in the protocol of sample preparation established for the identification of peptide biomarkers by HRMS.




| | Defatting | Extraction | purification |
|---|---|---|---|
| C1 | No | Sonnication | Desalt |
| C2 | No | Sonnication | Dilute |
| C3 | Yes | Sonnication | Desalt |
| C4 | Yes | Sonnication | Dilute |
| C5 | No | Overnight | Desalt |
| C6 | No | Overnight | Dilute |
| C7 | Yes | Overnight | Desalt |
| C8 | Yes | Overnight | Dilute |

*Figure 17 - Comparison of the impact of the extraction procedure (30 min agitation followed by 15 min sonication and 16 h agitation) and urea concentration reduction (desalting or dilution) on protein recovery of unprocessed hazelnut. Results are expressed as protein recovery and represent the means ± 1 S.D. for 3 independent experiments. Proteins were measured by monitoring the absorbance at 280 nm and protein recovery was calculated based on theoretical protein content of hazelnut (15 %).*




| | Defatting | Extraction | Urea reduction |
|---|---|---|---|
| C1 | No | Sonication | Desalt |
| C2 | No | Sonication | Dilute |
| C3 | Yes | Sonication | Desalt |
| C4 | Yes | Sonication | Dilute |
| C5 | No | Overnight | Desalt |
| C6 | No | Overnight | Dilute |
| C7 | Yes | Overnight | Desalt |
| C8 | Yes | Overnight | Dilute |

*Figure 18 - Comparison of the impact of the extraction procedure (30 min agitation followed by 15 min sonication and 16 h agitation) and urea concentration reduction (desalting or dilution) on the number of identified peptides for four hazelnut proteins (denominated with their UniProt identifier) from unprocessed hazelnut extract. Results are expressed as the number of identified peptides and represent the means ± 1 S.D. for 3 independent experiments. Statistical significance was evaluated with Student’s T-test (* = p < 0.05, ** = p < 0.01)*

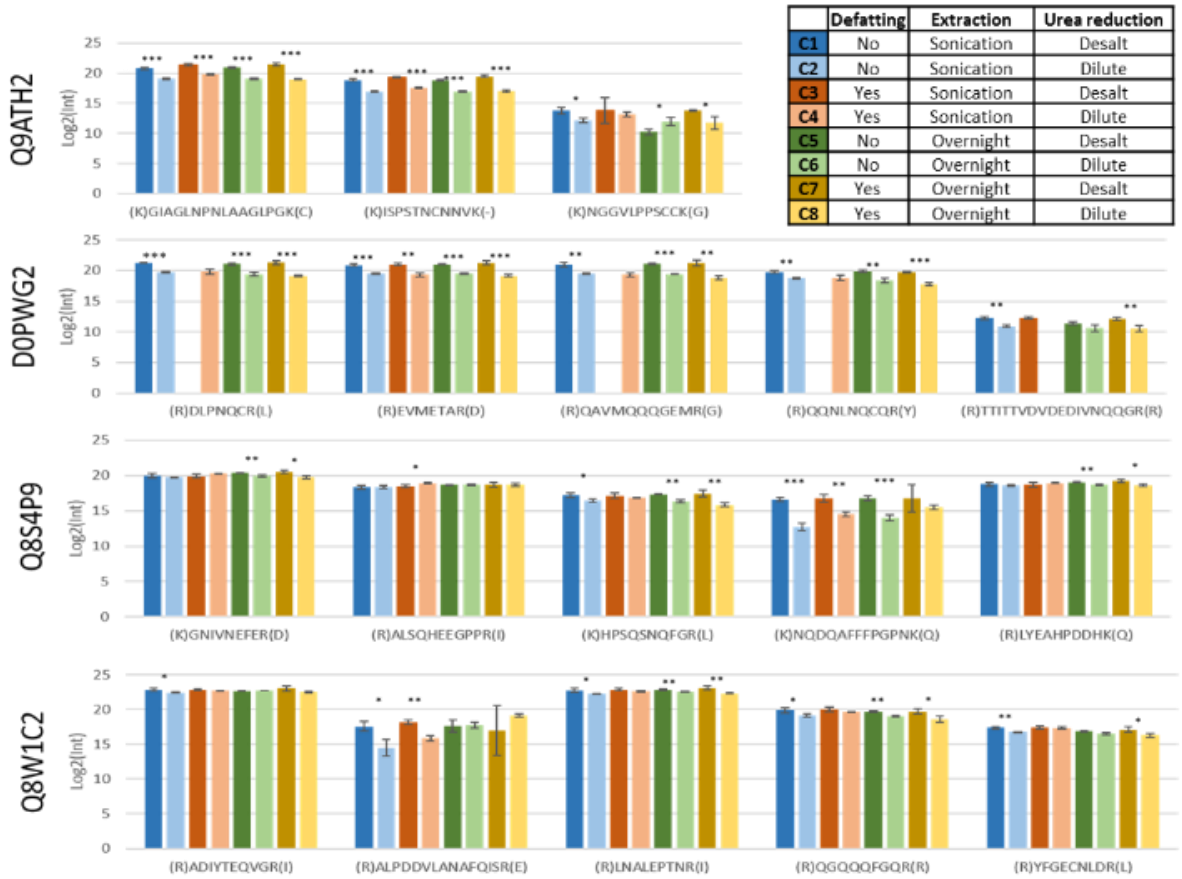



*Figure 19 - Comparison of the impact of the extraction procedure (30 min agitation followed by 15 min sonication and 16 h agitation) and urea concentration reduction (desalting or dilution) on MS signal intensity for different peptides of four hazelnut proteins (denominated with their UniProt identifier) from unprocessed hazelnut extract. Results are expressed as $Log_2$ of the MS signal intensity and represent the means ± 1 S.D. for 3 independent experiments. Statistical significance was evaluated with Student's T-test (* = p < 0.05, ** = p < 0.01, *** = p < 0.001)*

### 3.2. Optimization of the enzymatic digestion (by Maxime Gavage)

The enzymatic digestion of the extracted proteins is the second main step of sample preparation that needs to be optimized in the preparation of the sample. This step was optimized to achieve an optimal number of proteins and peptides that can be identified by HRMS. The decision on criteria to adopt/implement in the protocol was based on both the number of identified peptides (and thus proteins) and the MS signal intensity associated with the identified peptides.

The optimization of the digestion protocol was performed on unprocessed and roasted peanut and hazelnut extracts as the test matrices for these ingredients were the first to be generated (Figure 15). The main parameters to be optimized for the enzymatic digestion are the choice of the enzyme, the enzyme/protein ratio and the duration of the digestion.

The initial digestion protocol was:

- Protein extract desalted with 50 mM TetraEthylAmmonium Bicarbonate (TEAB); pH 9.2
- Protein concentration measurement based on absorbance at 280 nm
- Dilution of the sample to 1 mg/ml protein concentration in 50 mM TEAB
- Digestion (approximatively 16 h) at 37 °C under 300 rpm agitation with Trypsin Gold (Promega) (trypsin:protein ratio 1:20)

The digestion was stopped by the addition of trifluoroacetic acid (TFA, 1 % final concentration). The peptides resulting from the digestion were then purified using C18 solid phase extraction (SPE) cartridges (Waters) and analysed by HPLC-HRMS (maXis Impact UHR-TOF (Bruker) coupled to a Dionex UltiMate 3000 Standard LC Systems (Thermo Fisher Scientific) equipped with a reverse-phase liquid chromatography Acclaim PepMap100 C18 column (3 µm, 100 Å, 1 mm × 15 cm nanoViper, Thermo Fisher Scientific)). The same SPE purification protocol, HRMS analysis parameters and peptide identification criteria were used for each analysis and are extensively described in the different research articles exposed in section 4 (page 58): “Potential peptide biomarker identification”

Trypsin is well known for its highly specific activity (Olsen *et al*, 2004). Proteins are cleaved at the carboxy-terminal domain (C-term) after arginine and lysine except when these two amino acids are followed by a proline (Rodriguez *et al*, 2008). However, under the initial digestion conditions exposed above, it was observed that around 20 % of peptides obtained were not the result of a specific trypsin digestion (without arginine or lysine as C-term amino acid) and these were called “non-tryptic peptides”.

These non-tryptic peptides are a problem since they make the sample more complex and so, more difficult to identify by algorithm search. Moreover, non-tryptic peptides reduced the amounts of tryptic peptides reaching the MS detector, leading to a loss of sensitivity of the method.

Several hypotheses were then tested to determine the origin of the formation of “non-tryptic peptides”.

The first one was the putative presence of another digestive enzyme that could come from a contamination of the trypsin or the presence of enzymes from the sample itself. Mass spectrometry grade trypsin (Gold Trypsin from Promega) was used for robust peptide identification. Indeed, “low grade” trypsin is known for chymotrypsin contamination and thus, a part of the enzymatic activity is not related to the specific activity of trypsin (Burkhart *et al*, 2012). To evaluate this hypothesis, the relative cleavage frequency of every possible pair of amino acids was measured. This relative frequency was standardized by the frequency of each pair of amino acids in the identified sequences. With this approach, the presence of another enzyme, with a specific activity, could be identified. Chymotrypsin is for example known to preferentially cleave peptide bonds at C-term position of amino acids with large hydrophobic side chain such as tyrosine, tryptophan, and phenylalanine (Bender *et al*, 1973).

Results, gathering eight independent HRMS analyses of samples prepared with the initial digestion protocol described above, are presented in Table 4. The trypsin activity is clearly visible with a high relative cleavage frequency at the C-term position of lysine and arginine. No other specific cleavage was observed and so, no other specific enzymatic activity could be detected. Based on these results, the hypothesis supposing that trypsin could be contaminated with chymotrypsin was therefore rejected.

*Table 4 – Analysis of the relative cleavage frequency (%) of every possible pair of amino acids in order to identify the presence of a putative other digestive enzyme than trypsin with a different activity. The frequency of occurrence of each amino acid at N-term and C-term extremity of each identified peptide is normalized for the frequency of each pair of amino acids in the corresponding protein sequences.*

N-term amino acid

| C-term amino acid | A | C | D | E | F | G | H | I | K | L | M | N | P | Q | R | S | T | V | W | Y |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 4,0 | 2,6 | 1,6 | 1,0 | 0,2 | 0,7 | 0,0 | 1,9 | 0,0 | 1,0 | 7,4 | 1,5 | 0,5 | 1,0 | 7,6 | 0,6 | 4,7 | 3,0 | 0,0 | 0,0 |
| C | 5,6 | 0,0 | 0,0 | 0,0 | 13,3 | 3,8 | 0,0 | 39,1 | 9,1 | 3,2 | 0,0 | 0,0 | 0,0 | 1,4 | 0,0 | 0,0 | 0,0 | 5,1 | 0,0 | 0,0 |
| D | 0,7 | 0,0 | 1,3 | 3,2 | 2,2 | 7,0 | 5,6 | 5,6 | 1,0 | 1,5 | 0,0 | 1,8 | 1,5 | 0,4 | 1,8 | 4,3 | 3,0 | 2,9 | 16,7 | 0,0 |
| E | 1,3 | 1,4 | 1,6 | 0,8 | 0,7 | 0,6 | 0,0 | 1,8 | 0,8 | 1,6 | 0,0 | 0,4 | 0,5 | 1,2 | 0,0 | 1,5 | 1,2 | 3,1 | 2,3 | 0,0 |
| F | 14,8 | 4,2 | 0,5 | 0,0 | 2,8 | 8,1 | 0,0 | 8,0 | 2,4 | 1,7 | 6,7 | 4,8 | 0,0 | 6,2 | 0,0 | 13,5 | 5,5 | 1,8 | 50,0 | 15,3 |
| G | 1,4 | 6,8 | 1,0 | 2,0 | 1,0 | 3,3 | 0,0 | 4,0 | 5,3 | 2,2 | 5,4 | 5,5 | 0,0 | 0,5 | 3,8 | 5,0 | 4,7 | 0,6 | 3,5 | 4,1 |
| H | 4,4 | 42,9 | 0,0 | 2,0 | 11,1 | 7,3 | 10,5 | 14,7 | 33,3 | 7,0 | 60,0 | 0,0 | 0,0 | 6,0 | 7,9 | 28,2 | 10,5 | 17,0 | 10,0 | 14,3 |
| I | 0,4 | 0,0 | 0,4 | 0,3 | 0,2 | 0,7 | 0,0 | 2,3 | 0,0 | 1,2 | 0,0 | 0,7 | 5,3 | 0,5 | 0,7 | 0,0 | 0,6 | 0,4 | 0,0 | 0,4 |
| K | 92,6 | 73,1 | 66,7 | 84,7 | 88,0 | 83,1 | 97,8 | 94,5 | 85,8 | 88,2 | 90,9 | 80,2 | 10,3 | 89,6 | 83,9 | 93,4 | 93,8 | 89,0 | 87,5 | 96,0 |
| L | 1,5 | 7,9 | 0,8 | 0,8 | 1,0 | 5,3 | 0,0 | 6,2 | 1,0 | 2,8 | 1,8 | 2,2 | 1,8 | 0,7 | 0,9 | 8,1 | 5,3 | 1,6 | 25,0 | 9,6 |
| M | 12,5 | 0,0 | 0,0 | 1,1 | 3,3 | 5,3 | 0,0 | 9,4 | 9,1 | 4,7 | 2,6 | 2,0 | 0,0 | 0,0 | 3,7 | 0,0 | 10,0 | 5,4 | 0,0 | 25,6 |
| N | 7,4 | 13,3 | 11,2 | 5,9 | 1,7 | 39,0 | 4,2 | 8,6 | 8,1 | 5,2 | 11,8 | 7,7 | 3,0 | 8,0 | 0,7 | 14,6 | 17,9 | 7,2 | 5,9 | 7,0 |
| P | 1,0 | 8,7 | 0,0 | 0,2 | 2,8 | 1,6 | 0,0 | 4,1 | 0,0 | 4,8 | 0,0 | 2,0 | 0,8 | 0,0 | 0,3 | 3,5 | 0,0 | 4,9 | 0,0 | 0,0 |
| Q | 3,8 | 0,7 | 0,8 | 0,2 | 0,7 | 3,8 | 0,7 | 2,5 | 0,0 | 2,0 | 2,3 | 6,8 | 1,3 | 3,6 | 1,2 | 4,9 | 3,7 | 0,4 | 5,9 | 1,4 |
| R | 89,2 | 87,6 | 90,7 | 91,3 | 76,7 | 90,8 | 93,5 | 97,7 | 87,7 | 91,5 | 85,7 | 96,3 | 31,9 | 96,8 | 93,1 | 95,9 | 94,4 | 94,3 | 100,0 | 98,3 |
| S | 1,3 | 0,0 | 1,3 | 4,8 | 1,2 | 3,1 | 0,0 | 5,7 | 0,8 | 1,1 | 7,7 | 2,3 | 1,4 | 1,9 | 1,0 | 3,5 | 5,0 | 1,8 | 2,3 | 0,4 |
| T | 5,0 | 2,7 | 0,8 | 1,2 | 0,9 | 4,9 | 16,7 | 3,1 | 1,7 | 3,7 | 0,0 | 0,0 | 0,8 | 1,4 | 0,0 | 4,7 | 5,0 | 0,4 | 0,9 | 0,7 |
| V | 1,4 | 0,0 | 0,3 | 0,8 | 1,6 | 1,7 | 3,0 | 0,6 | 0,0 | 1,5 | 10,7 | 0,0 | 1,0 | 1,0 | 0,0 | 0,0 | 2,0 | 0,5 | 0,0 | 0,0 |
| W | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 3,4 | 0,0 | 6,7 | 0,0 | 12,1 | 22,0 | 0,0 | 0,0 | 0,0 | 0,0 | 6,3 | 0,0 | 2,9 | 0,0 | 0,0 |
| Y | 4,9 | 10,0 | 0,0 | 0,0 | 5,5 | 0,9 | 0,0 | 9,8 | 0,0 | 6,6 | 15,4 | 2,7 | 0,0 | 2,3 | 1,5 | 9,9 | 13,7 | 4,1 | 0,0 | 0,0 |

A second hypothesis we tested was the potential fragmentation of the peptides in the ionisation source. Indeed, the energy transmitted to the peptides can be sufficient to induce the break of peptide bonds and so, could generate non-tryptic peptides (Kim *et al*, 2013). This kind of fragmentation can be identified by the retention time of the peptides coming from a same parent peptide. Indeed, since the parent peptide is fragmented in the ionisation source, the fragmented peptides reach the mass spectrometer detector at the same time (neglecting the time of flight in the mass spectrometer). So, peptides coming from the same parent peptide are characterised by the same retention time in the liquid chromatography column (Kim *et al*, 2013).

In this case, observed retention times for peptides coming from the same parent peptides were variable and so, the in-source fragmentation was excluded. To illustrate this, a relevant example of a peptide from Arachin Ahy-3 peanut protein is presented in Figure 20.

Another hypothesis was the presence of pseudo-trypsin. Pseudo-trypsin or ψ-trypsin has a less specific activity when compared to trypsin and is a product of trypsin autolysis by the cleavage of the $Lys_{176}$-$Asp_{177}$ bond (Perutka & Sebela, 2018). The "Promega Gold Trypsin – Mass Spectrometry Grade", used for the enzymatic digestion is supposed to be protected against autolysis by a reductive methylation of lysine side-chain. However, if the reductive methylation is not total, autolysis can occur (Rice *et al*, 1977). Trypsin autolysis can be modelled as a second order kinetic rate reaction (Sriram *et al*, 1996). In other words, a longer digestion duration or a higher trypsin concentration promote ψ-trypsin formation and therefore non-tryptic peptides generation.

| Retention time (min) | Peptide sequence |
|---|---|
| 25,76 | I F P G C P S T Y E E P A Q E G R |
| 31,26 | R P F Y S N A P L E I Y |
| 31,97 | R P F Y S N A P L E I Y V Q Q G |
| 32,01 | Y S N A P L E I Y |
| 32,01 | R P F Y S N A P L E I Y V Q Q G S |
| 33,14 | R P F Y S N A P L E I Y V Q Q G S G Y |
| 33,75 | F G L I F P G C P S T Y E E P A Q E G R |
| 35,96 | V Q Q G S G Y F G L I F P G C P S T Y E E P A Q E G R |
| 36,11 | R P F Y S N A P L E I Y V Q Q G S G Y F |
| 38,23 | Y S N A P L E I Y V Q Q G S G Y F |
| 38,53 | R P F Y S N A P L E I Y V Q Q G S G Y F G L |
| 39,07 | R P F Y S N A P L E I Y V Q Q G S G Y F G L I F P G C P S T Y E E P A Q E G R |
| 40,61 | S N A P L E I Y V Q Q G S G Y F G L I F P G C P S T Y E E P A Q E G R |
| 41,11 | Y S N A P L E I Y V Q Q G S G Y F G L I F P G C P S T Y E E P A Q E G R |
| 41,46 | Y S N A P L E I Y V Q Q G S G Y F G L |

*Figure 20 – Comparison of the liquid chromatography retention times of non-tryptic peptides coming from the same parent peptide, highlighted in colour (peptide from Arachin Ahy-3, a peanut protein).*

To test this hypothesis, different times of digestion (2, 5 and 16 h) and two trypsin:protein ratios (1:20 and 1:50) were tested (Figure 21). The digestion conditions influenced the proportion of non-tryptic peptides in identified peptides. A longer digestion duration or a higher enzyme:protein ratio significantly (in most of the cases) increased the proportion of non-tryptic peptides. As expected, the proportion of non-tryptic peptides negatively influences the identification of tryptic peptides and thus proteins (Figure 22).

In conclusion for this non-specific digestion problematic, the solution to reduce the proportion of non-tryptic peptides and obtain an optimal number of protein and tryptic peptide identifications was to use a short digestion time or a low trypsin concentration. This second strategy was implemented and we adopted, for practical reasons, an overnight digestion (16 h) and a 1:50 enzyme:protein ratio.

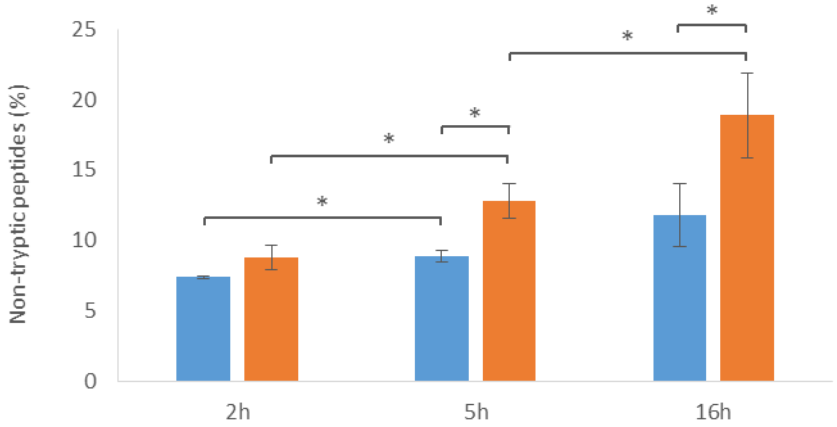



*Figure 21 – Effects of digestion conditions (incubation time with trypsin and enzyme:protein ratio (1:50 in blue and 1:20 in orange)) on the proportion of non-tryptic peptides in identified peptides in unprocessed peanut samples. Results are expressed as the proportion of non-tryptic peptides in identified peptides and represent the means ± 1 S.D. for 3 independent experiments (separate digestions and HRMS analysis of the same unprocessed peanut extract). Statistical significance was evaluated with Student's T-test (* = $p < 0.05$)*

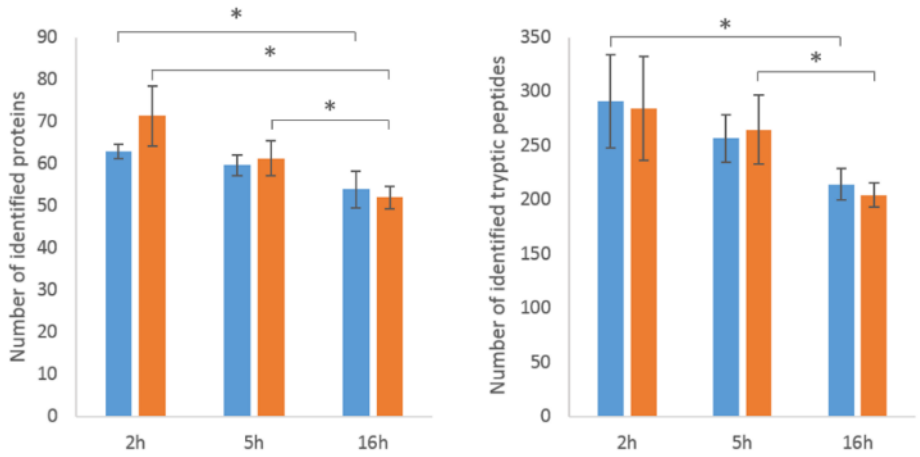
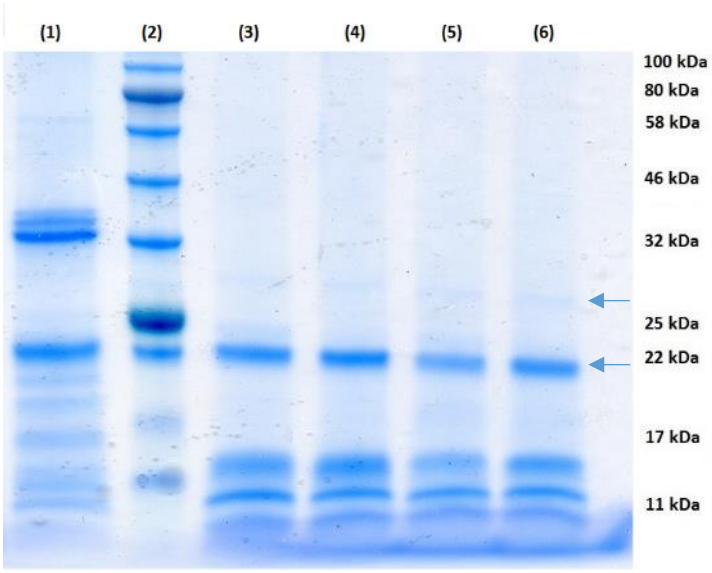



*Figure 22 – Effects of digestion conditions (incubation time with trypsin and enzyme:protein ratio (1:50 in blue and 1:20 in orange)) on the number of identified proteins and peptides in unprocessed peanut samples. Results are expressed as the number of identified proteins and peptides and represent the means ± 1 S.D. for 3 independent experiments (separate digestions and HRMS analysis of the same unprocessed peanut extract). Statistical significance was evaluated with Student's T-test (* = $p < 0.05$)*

Effects of digestion conditions were also observed on the completeness of protein digestion.

As we can observe in Figure 23, some proteins are resistant to trypsin digestion. Unprocessed peanut protein extract was digested under different conditions (several digestion times and two different enzyme:protein ratios) and their abundance was analysed by sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) and Coomassie blue staining. One can observe that no matter the digestion duration (2 or 16h) or the enzyme:protein ratio (1:20 or 1:50), some bands, corresponding to undigested proteins or part of proteins remained present/intact (indicated with arrows on Figure 23).




Caption:

(1): Undigested extract

(2): Molecular weight marker

(3): 2 h digestion (1:20)

(4): 2 h digestion (1:50)

(5): 16 h digestion (1:20)

(6): 16 h digestion (1:50)

*Figure 23 – SDS-PAGE with Coomassie blue staining analysis of the different tryptic digestion conditions (digestion time of 2 or 16 h and enzyme:protein ratios 1:20 and 1:50) of unprocessed peanut protein extract. Samples corresponding to 5 µg of proteins were loaded on an SDS-PAGE gel (12 %) and the electrophoresis was run under 150 mV.*

Peanuts are composed of 25 % of proteins and some of them such as conglutins, are well known, due to their conformational stability, to be resistant to digestive enzymes like trypsin (Apostolovic *et al*, 2016). The resistance to enzymatic digestion is an issue in the development of a method to quantify allergens because it reduces the number of potential peptide biomarkers and thus introduces a bias in the quantification.

Different approaches were then tested to improve the digestion yield.

Digestion resistant proteins were first identified through the combination of SDS-PAGE resolution and HRMS. Bands corresponding to trypsin digestion resistant proteins were cut out of the gel and placed in separated containers. Protein contained in these bands were then digested within the gel (see Annex 1 on page 213 for UNamur in-gel protein digestion protocol) and the obtained peptides were analysed by HRMS. The second digestion is supposed to be more efficient as the proteins are totally denatured in the gel (Huynh *et al*, 2009). This procedure was applied to digestion resistant proteins observed on SDS-PAGE analysis of trypsin digested unprocessed peanut extract.

As observed in Figure 23, for peanut protein sample digestion, two bands corresponding to digestion resistant proteins were cut out of the gel and analysed by HRMS. The first one corresponds to the major resistant band observed in trypsin-digested extract with a molecular weight around 22 kDa. The second one was a less abundant fragment band characterized by a molecular weight around 30 kDa. These two bands are indicated with an arrow on Figure 23.

The 30 kDa band was identified as a galactose-binding lectin (P02872 Uniprot identifier). According to identification acceptance criteria detailed in the published research articles presented in section 4 (page 58), the 5 identified peptides corresponded to this protein (Figure 24). It is known that lectins are described in the literature as proteins that are resistant to heat and digestive enzymes (Romano & Castagna, 2016).

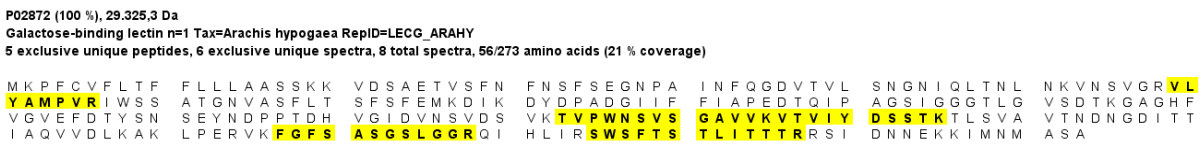



*Figure 24 – HRMS analysis of the 30 kDa trypsin digestion resistant protein band. Identified peptides are highlighted in yellow in the peanut galactose binding lectin (P02872 UniProt identifier) protein sequence. Scaffold proteomic software was used to analyse and visualise HRMS analysis results* (Searle, 2010).

The 22 kDa band was identified as a part of Ara h 3, a major peanut allergen (Koppelman *et al*, 2010). This seed storage protein is composed of 2 subunits linked by a disulphide bond (cleaved during the reduction/alkylation reaction preceding the enzymatic digestion) (Koppelman *et al*, 2003). The first subunit is located at the N-terminus of the protein sequence and is identified as “acidic subunit” with a molecular weight of 45 kDa. The second one is located at the C-terminus of the protein sequence and is identified as “basic subunit” with a molecular weight of 23 kDa (Figure 25).

>tr|Q0GM57|Q0GM57_ARAHY Iso-Ara h3 OS=Arachis hypogaea PE=2 SV=1

MAKLLALSLCFCVLVLGASSVTFRQGGEENECQFQRLNAQRPDNRIESEGGYIETWNPNN
QEFQCAGVALSRTVLRRNALRRPFYSNAPLEIYVQQGSGYFGLIFPGCPSTYEEPAQEGR
RYQSQKPSRRFQVGQDDPSQQQQDSHQKVHRFDEGDLIAVPTGVAFWMYNDEDTDVVTVT
LSDTSSIHNQLDQFPRRFYLAGNQEQEFLRYQQQQGSRPHYRQISPRVRGDEQENEGSNI
FSGFAQEFLQHAFQVDRQTVENLRGENEREEQGAIVTVKGGLRILSPDEEDESSRSPPSR
REEFDEDRSRPQQRGKYDENRRGYKNGIEETICSASVKKNLGRSSNPDIYNPQAGSLRSV
NELDLPILGWLGLSAQHGTIYRNAMFVPHYTLNAHTIVVALNGRAHVQVVDSNGNRVYDE
ELQEGHVLVVPQNFAVAAKAQSENYEYLAFKTDSRPSIANLAGENSIIDNLPEEVVANSY
RLPREQARQLKNNNPFKFFVPPFDHQSMREVA

Acidic subunit - Basic subunit

*Figure 25 – Amino acid sequence of the two subunits (acidic and basic) in the Ara h 3 protein sequence (from Scaffold software)*

In addition, nine out of the ten identified peptides for the 22 kDa band were situated in the basic subunit of Ara h 3 protein sequence (Figure 26). Moreover, the molecular weight of the band on the gel and the theoretical/predicted molecular weight of the basic subunit are comparable.

**Q0GM57 (100 %), 58.262,7 Da**
**Iso-Ara h3 n=2 Tax=Arachis hypogaea RepID=Q0GM57_ARAHY**
**10 exclusive unique peptides, 21 exclusive unique spectra, 91 total spectra, 177/512 amino acids (35 % coverage)**

| | | | | | | |
|---|---|---|---|---|---|---|
| MAKLLALSLC | FCVLVLGASS | VTFRQGGEEN | ECQFQRLNAQ | RPDNRIESEG | GYIETWNPNN | QEFQCAGVAL |
| SRTVLRRNAL | RRPFYSNAPL | EIYVQQGSGY | FGLIFPGCPS | TYEEPAQEGR | RYQSQKPSRR | FQVGQDDPSQ |
| QQQDSHQKVH | RFDEGDLIAV | PTGVAFWMYN | DEDTDVVTVT | LSDTSSIHNQ | LDQFPRRFYL | AGNQEQEFLR |
| YQQQQGSRPH | YRQISPRVRG | DEQENEGSNI | FSGFAQEFLQ | HAFQVDRQTV | ENLRGENERE | EQGAIVTVKG |
| GLRILSPDEE | DESSRSPPSR | REEFDEDRSR | PQQRGKYDEN | RRGYKNGIEE | TICSASVKKN | LGRSSNPDIY |
| NPQAGSLRSV | NELDLPILGW | LGLSAQHGTI | YRNAMFVPHY | TLNAHTIVVA | LNGRAHVQVV | DSNGNRVYDE |
| ELQEGHVLVV | PQNFAVAAKA | QSENYEYLAF | KTDSRPSIAN | LAGENSIIDN | LPEEVVANSY | RLPREQARQL |
| KNNNPFKFFV | PPFDHQSMRE | VA | | | | |

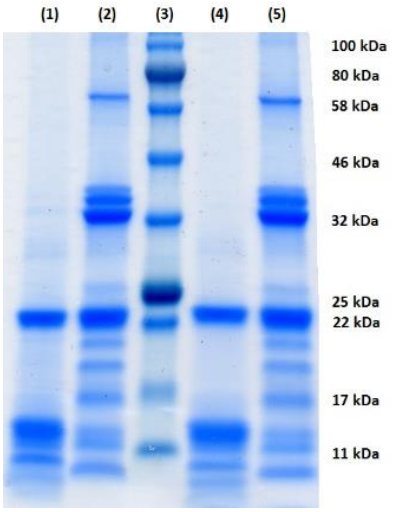
*Figure 26 – HRMS analysis of the 22 kDa trypsin digestion resistant protein band. Identified peptides are highlighted in yellow in the peanut Ara h 3 (Q0GM57 UniProt identifier) protein sequence (from Scaffold software). Amino acids highlighted in green were found as modified amino acids (methionine oxidation and cysteine carbamidomethylation)*

As we have seen, with the combined approach of SDS-PAGE resolution and HRMS analysis, peanut Ara h 3 and galactose-binding lectin were identified as proteins that are resistant to trypsin digestion and were considered as good markers for the optimization of enzymatic digestion.

Resistance to enzymatic digestion might be due to the protein conformation leading to inaccessible cleavage sites (Koppelman *et al*, 2010). The putative effect of protein denaturation was therefore tested in order to improve the enzymatic digestion of these resistant proteins.

The effect of a mass spectrometry friendly ionic detergent such as sodium deoxycholate (SDC), was first analysed. The use of SDC for in-solution digestion is indeed described in proteomic literature to enhance the activity of trypsin and increase the recovery of hydrophobic peptides and the number of identified proteins. Wang and co-workers demonstrated that the use of SDC for in-solution digestion had a positive effect on the detection of membrane proteins of barley leaf (Wang *et al*, 2018). SDC is also cheaper than RapiGest, a characteristic which has to be taken into account in the perspective of the development of a routine method.

The effect of 1 % SDC, an optimal concentration according to Lin and co-workers (Lin *et al*, 2008), on the digestion was evaluated on raw peanut protein extracts. As presented in Figure 27, the presence of 1 % SDC does not strongly improve the digestion of resistant proteins by trypsin. The 22 kDa band was present with comparable intensities between samples digested by trypsin in the presence or not of 1 % SDC.




Caption:

(1): 2 h digestion (1:20 – 0 % SDC)

(2): Undigested extract (0 % SDC)

(3): Molecular weight marker

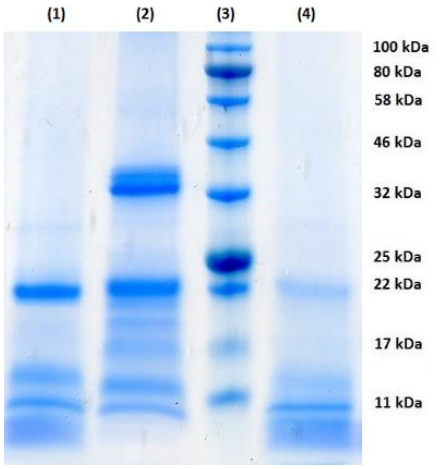
(4): 2 h digestion (1:20 – 1 % SDC)

(5): Undigested extract (1 % SDC)

*Figure 27 – Evaluation of the effect of 1 % SDC on the trypsin digestion efficiency of unprocessed peanut protein extracts. Samples corresponding to 5 µg of proteins were loaded on an SDS-PAGE gel (12 %), the electrophoresis was run under 150 mV and the gel stained with Coomassie blue.*

The second strategy we tested to improve protein digestion was the sequential combination of two different enzymes (Endoproteinase Lys-C and trypsin) and the use of a higher urea concentration. Endoproteinase Lys-C is a serine endoproteinase which cleaves peptide bonds at the C-term part after lysine residues (Raijmakers *et al*, 2010). Its specificity is therefore compatible with trypsin. However, unlike trypsin, Lys-C is still active under high urea concentrations (6−8 M) (Glatter *et al*, 2012). To improve protein digestion, a two-step digestion was evaluated.

The use of a gel filtration column before the enzymatic digestion step (see protein extraction optimization section above, Paragraph 3.1 page 46), allowed the exchange of the protein extract buffer for 50 mM TEAB containing 6 M urea. According to Lys-C provider recommendations (Wako Pure Chemical Industries, Osaka, Japan), a Lys-C pre-digestion step was carried out during 5 h at 37 °C with a 1:50 enzyme:protein ratio. Samples were then diluted with 50 mM TEAB to reduce urea concentration to 1 M, a concentration that is compatible with trypsin activity (according to Promega). Enzymatic digestion was pursued with the addition of trypsin (1:20) and an incubation of 2 h at 37°C. The complete/detailed protocol can be found in the published research articles exposed in section 4 (page 58).




Caption:

(1): Trypsin digestion (2h – 1:20)

(2): Undigested extract

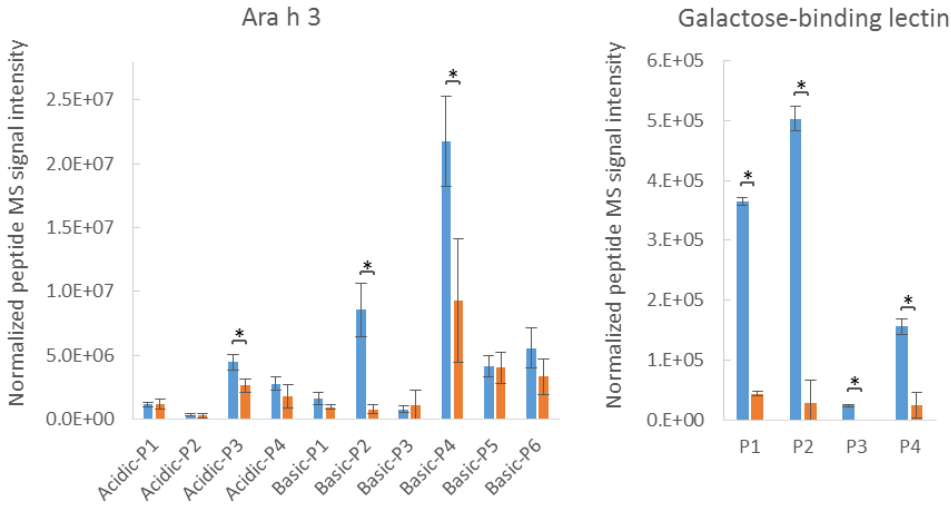
(3): Molecular weight marker

(4): Tandem Lys-C (5h – 1:50) and trypsin (2h – 1:20) digestion

(Normalized protein loading)

*Figure 28 – Effects of the addition of a Lys-C pre-digestion step (5 h at 37°C in 6 M urea) on the trypsin digestion of unprocessed peanut protein extracts. Samples corresponding to 5 µg of proteins were loaded on a SDS-PAGE gel (12 %), the electrophoresis was run under 150 mV and the gel stained with Coomassie blue.*

The effect of the digestion with the sequential combination of Lys-C/trypsin was compared to the digestion efficacy by trypsin alone (digestion time: 2 h – enzyme:protein ratio 1:20). Digested products were analysed with SDS-PAGE stained with Coomassie blue and by HRMS analysis a particular focus on peptides obtained from the digestion of resistant proteins (Ara h 3 and galactose-binding lectin) (Figure 28). The results show that the Lys-C/trypsin digestion is able to improve the trypsin digestion for resistant proteins when compared to the digestion by trypsin alone. Next, the staining intensity of the bands observed on the SDS gel was also analysed with ImageJ, an image analysis software. This analysis revealed that the 22 kDa band intensity was reduced by 80 % for the sample digested by the combination of Lys-C and trypsin when compared to the sample digested by trypsin only.

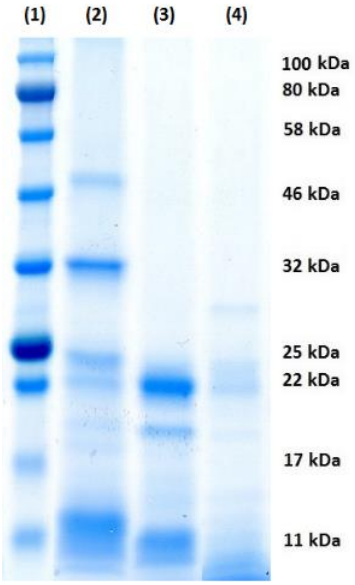



| Ara h 3 acidic subunit | Ara h 3 basic subunit | Galactose-binding lectin |
|---|---|---|
| P1: FYLAGNQEQEFLR | P1: AQSENYEYLAFK | P1: VLYAMPVR |
| P2:GDEQENEGSNIFSGFAQEFLQHAFQVDR | P2: NAMFVPHYTLNAHTIVVALNGR | P2: DYDPADGIIFFIAPEDTQIPA GSIGGGTLGVSDTK |
| P3:IESEGGYIETWNPNNQEFQCAGVALSR | P3: SSNPDIYNPQAGSLR | P3: TLSVAVTNENGDITTIAQVVDLK |
| P4: RPFYSNAPLEIYVQQGSGYFGL IFPGCPSTYEEPAQEGR | P4: SVNELDLPILGWLGLSAQHGTIYR | P4: SWSFTSTLITTTR |
| | P5: TDSRPSIANLAGENSIID NLPEEVVANSYR | |
| | P6: VYDEELQEGHVLVVPQNFAVAAK | |

*Figure 29 – HRMS analysis of the effect the two digestion protocols (sequential Lys-C/trypsin treatment (in blue) and trypsin alone (in orange)) applied on unprocessed peanut extract on the digestion efficiency. Two resistant protein markers (Ara h 3 and galactose-binding lectin) are considered. Peptides of Ara h 3 are separated between acidic and basic subunit. Results are expressed as the peptide MS signal intensity and represent the means ± 1 S.D. for 3 independent experiments. Statistical significance was evaluated with Student's T-test (* = $p < 0.05$).*

These results were confirmed by the HRMS analysis. As presented in Figure 29, the signal intensity associated with peptides from trypsin-resistant proteins was higher in samples digested with the two enzymes. This observation is particularly true for the digestion of the galactose-binding lectin with a signal intensity that is statistically significantly higher for the four identified peptides obtained when the samples are digested by both enzymes when compared to the digestion by trypsin only. Results are less clear for Ara h 3 and its two subunits, as the signal intensity is only statistically significantly higher for a few peptides when the sample was digested by both enzymes.

The comparison of the two digestion protocols was also performed on raw hazelnut protein extracts (Figure 30). One can observe that the digestion of trypsin resistant proteins is also improved with the use of the sequential Lys-C/trypsin digestion protocol.




Caption:

(1): Molecular weight marker

(2): Undigested extract

(3): Trypsin digestion (2h – 1:20)

(4): Tandem Lys-C (5h – 1:50) and trypsin (2h – 1:20) digestion

(Normalized protein loading)

*Figure 30 - Effects of the addition of a Lys-C pre-digestion step (5 h at 37°C in 6 M urea) on the trypsin digestion of unprocessed hazelnut protein extract. Samples corresponding to 5 µg of proteins were loaded on a SDS-PAGE (12 %), run under 150 mV and stained with Coomassie blue.*

In conclusion, the combined optimisation of protein extraction and enzymatic digestion conditions resulted in a complete sample preparation protocol presented in Figure 31. Both digestion protocols were applied in parallel to each test matrix to obtain a more complete peptide identification and allow the identification of proteins that are resistant to digestion by trypsin alone.

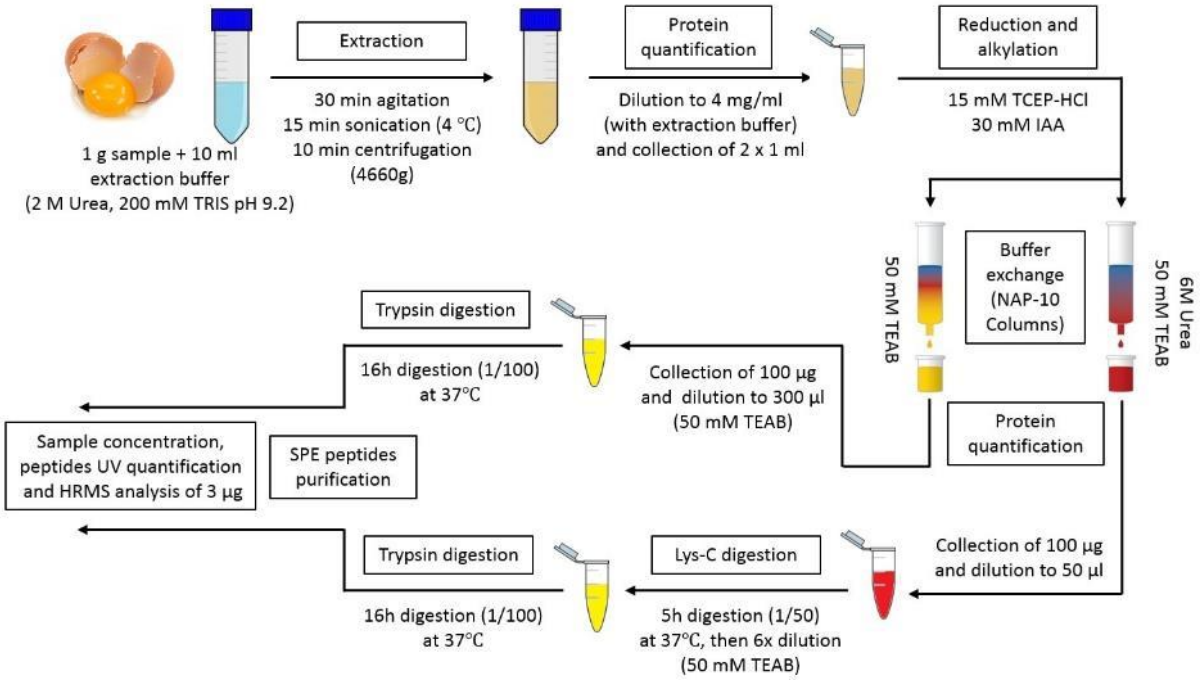



*Figure 31 - Sample preparation flowchart. Proteins from a 1 g sample were extracted with 10 mL extraction buffer. The proteins were then reduced and alkylated. The buffer was exchanged with an appropriate buffer depending on the subsequent proteolytic digestion (trypsin alone or tandem Lys-C/trypsin). The resulting peptides were purified with the SPE before HRMS analysis was performed.*

## 4. Potential peptide biomarker identification

The developed and optimized sample preparation protocol was systematically applied to the different test matrices produced for each of the four allergenic ingredients. Resulting samples were analysed by HRMS to identify potential peptide biomarkers for the quantification of food allergens in processed food products.

The obtained results are now presented in the different published articles or in manuscript in preparation for the different allergenic ingredients tested. The article on eggs was published in *Journal of Chromatography A*. The articles on peanuts and hazelnuts were published in *Food Chemistry* Journal. At the time of writing this thesis, the fourth and last manuscript focusing on milk peptides was submitted to *Food Analytical Methods* Journal.

### 4.1. "Selection of egg peptide biomarkers in processed food products by high resolution mass spectrometry" (Gavage *et al*, 2019)

The developed peptide biomarkers identification strategy was first applied to egg. Egg allergy is one of the most common food allergies, especially in children (Eggesbø *et al*, 2001; Sicherer & Sampson, 2006). In addition, egg and its derived components are abundantly used by the food industry, thus making the complete avoidance of eggs difficult.

The developed and optimized sample preparation protocol was applied to processed egg matrices: freeze-dried whole liquid egg as unprocessed matrix, freeze-dried pasteurized whole liquid egg as heated matrix, mayonnaise as fat-rich matrix, mayonnaise containing vinegar as a fat-rich and low pH matrix and chocolate as a complex and fat-rich matrix. Resulting peptides were analysed by HPLC-HRMS.

The identified peptides were filtered using a series of criteria to ensure method specificity, sensitivity and robustness. To summarize, peptides must be specific for the allergenic ingredient, be robust to food processing, have originated from abundant proteins, be size-compatible with triple quadrupole mass spectrometry analysis, and not be prone to amino acid modifications or missed cleavages by the protease(s) used. The use of a tandem Lys-C/trypsin enzymatic digestion improved the digestion of some of the proteins as compared to the digestion with trypsin alone.

Considering these criteria, a list of 16 robust potential peptide biomarkers was obtained, allowing for the detection of the egg yolk and egg white proteins for eggs from hens and other farm birds (goose, duck, turkey, quail, and guinea fowl).

The 16 identified potential peptide biomarkers were considered for the development of the quantitative UHPLC-MS/MS analysis method.

The research article was published in *Journal of Chromatography A*.

Journal of Chromatography A, 1584 (2019) 115–125

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

# Selection of egg peptide biomarkers in processed food products by high resolution mass spectrometry

M. Gavage[a,b,*], K. Van Vlierberghe[c], C. Van Poucke[c], M. De Loose[c], K. Gevaert[d,e], M. Dieu[b,f], P. Renard[b], T. Arnould[b], N. Gillard[a]

[a] CER Groupe, Rue du Point du Jour 8, 6900, Marloie, Belgium
[b] Laboratory of Biochemistry and Cell Biology (URBC)-Namur Research Institute for Life Sciences (NARILIS), University of Namur, 61, Rue de Bruxelles, 5000, Namur, Belgium
[c] Flanders Research Institute for Agriculture, Fisheries and Food (ILVO), Brusselsesteenweg 370, 9090, Melle, Belgium
[d] VIB Center for Medical Biotechnology, Albert Baertsoenkaai 3, B9000, Ghent, Belgium
[e] Department of Biochemistry, Ghent University, Albert Baertsoenkaai 3, B9000, Ghent, Belgium
[f] MaSUN, Mass Spectrometry Facility, University of Namur, 61, Rue de Bruxelles, 5000, Namur, Belgium

ARTICLE INFO

*Article history:*
Received 3 August 2018
Received in revised form 5 November 2018
Accepted 18 November 2018
Available online 20 November 2018

*Keywords:*
High-resolution mass spectrometry
Allergen
Egg
Processed food products
Peptide biomarker selection

ABSTRACT

Food allergy is a growing health problem worldwide; thus, there is an urgent need for robust, specific, and sensitive analytical methods for detecting allergens. Mass spectrometry is an alternative to the existing methods, and it can overcome their limitations. One of the first steps in the development of any analytical method is the identification of the analytes to be further studied. In the case of allergen detection by mass spectrometry, the analytes are peptides. In this study, a strategy was developed for identifying potential peptide biomarkers in processed food products. This strategy was applied to processed egg matrices, and 16 potential peptide biomarkers were identified for the further detection and quantification of egg by means of mass spectrometry. With an empirical approach based on dedicated sample preparation, including tandem Lys-C/trypsin enzymatic digestion and high-resolution mass spectrometry analysis, hundreds of peptides from egg proteins were identified. This list of peptides was further refined with a series of criteria, obtained from empirical evidence, to identify the ideal biomarkers for the development of a quantitative method. These criteria include the resistance to food processing and the specificity of the peptides for eggs but also the effects of amino acid modifications and enzymatic digestion efficiency.


## 1. Introduction

Food allergy is defined by the National Institute of Allergy and Infectious Diseases (NIAID) as “an adverse health effect arising from a specific immune response that occurs reproducibly on exposure to a given food.” It is a major global issue, and children, in particular, are being affected in increasing numbers [1]. Symptoms can affect multiple biological systems like the skin (hives), or the gastrointestinal (vomiting and/or stomach cramps) or respiratory (wheezing, hoarse throat) systems. They can become life-threatening in the case of anaphylaxis simultaneously affecting different systems [2]. Because there is no cure yet, the only solution for allergic patients is strict avoidance of exposure to food allergens [3]. To protect consumers, 14 ingredients that have the potential to cause allergic reactions are required by European legislation (Regulation 1169/2011/CE) to be included in food labelling. Unfortunately, the incidental presence of allergens in food is possible, e.g., through cross-contamination, and this is not covered by legislation. Hence, specific and sensitive methods for the routine detection of food allergens are needed.

Some detection methods are commercially available, and they are based mainly on immunological methods targeting specific proteins, such as enzyme-linked immunosorbent assays (ELISA). Other methods rely on polymerase chain reactions (PCRs) targeting specific DNA molecules. However, such methods have limitations. For instance, with the PCR method, milk cannot be distinguished from cattle products nor egg from other chicken parts because they share the same DNA sequence. Furthermore, while the DNA is targeted, the proposed thresholds are expressed as total protein quantities; thus, the identification of the required conversion factor could be challenging [4]. The antibodies used in immunological methods recognize epitopes. One drawback, however, is the possible recognition by the antibodies of similar structures in non-allergenic ingredients, thus generating false positive results.

* Corresponding author at: CER Groupe, Rue du Point du Jour, 8, 6900, Marloie, Belgium.
*E-mail address:* n.gillard@cergroupe.be (M. Gavage).

https://doi.org/10.1016/j.chroma.2018.11.036
0021-9673/

In contrast, protein denaturation upon food processing can modify the conformation of the targeted proteins or epitopes, leading to false negative results [5].

Mass spectrometry (MS)-based methods may overcome these drawbacks. Typically, proteins are detected and quantified through their constitutive peptides, which are obtained after proteolytic digestion. Such an analysis is therefore based on peptide sequences rather than protein structures. Thus, MS-based methods are more robust than other methods. However, MS may be subject to matrix effects, defined as the combined effects of all the components of the sample other than the analyte on the measurement of the quantity of this specific analyte [6]. These matrix effects have to be minimized by dedicated sample preparation. The overall specificity of MS-based methods is ensured by the specificity of both the selected peptides and the multiple-reaction monitoring (MRM) transitions [7]. In addition, MRM is highly sensitive, and it is compatible with the detection of multiple analytes in a single experiment. Therefore, with MRM, multiple allergenic ingredients can be detected and quantified in a single analysis for a particular sample [8,9].

As was mentioned, with MS-based methods, the detection and the quantification of allergenic ingredients are performed by analyzing the constitutive peptides of the proteins from these food ingredients. The selection of these peptides is the first step in the development of the analytical method. This can be done with an *in silico* approach using software, e.g., Skyline®, to generate MRM transitions that will be assessed by analyzing the samples with a triple quadrupole mass spectrometer (QqQ) [10].

In this study, peptides were selected with an empirical approach based on high-resolution mass spectrometry (HRMS). Such an untargeted approach allowed for the selection of the peptides, taking into account peptide-related information such as natural and artificial amino acid modifications, and proteolytic digestion efficiency through the detection of incompletely digested peptides. This important information was not readily available with the *in silico* approaches mentioned above. In practice, proteins are extracted from matrices that may or may not be contaminated with allergenic material. After proteolytic digestion, the resulting peptides are analyzed by HRMS. The obtained data are then used to select peptide biomarkers for the development of the actual quantitative detection method. To achieve this goal, a digestion protocol combining two enzymes, Lys-C and trypsin, was implemented because it was observed in preliminary tests that digestion using only trypsin was inefficient for some egg proteins.

It is clear that peptide selection is a crucial step in the development of a quantitative MS-based method. These peptides have to fulfil a series of criteria to ensure method accuracy, robustness, and precision [11–13]. Ideal peptide biomarkers have to be specific for the allergenic ingredient, robust to food processing, derivable from abundant proteins, size-compatible with *m/z*-analyzers (8–25 amino acids), and not prone to amino acid modifications or missed cleavages by the protease(s) used.

This study is part of a larger project aimed at developing MS-based methods for the quantification of four main allergens—egg, milk, peanut and hazelnut—in processed food products. This particular study was focused on egg, but the developed strategy should be applicable for detecting any allergenic ingredient. However, this strategy depends on the availability of protein databases. This could be a limitation for less studied allergenic ingredients like pecan (*Carya illinoinensis*) with only 45 entries in UniProt protein database, for the moment.

Egg allergy is one of the most common food allergies, especially in children [14,15]. In addition, egg and its derived components are used a great deal by the food industry, thus making the complete avoidance of eggs difficult. To identify robust and specific peptides for egg, egg-based reference materials were prepared in a standardized way. The analytical criteria were also identified and applied to the final selection of egg peptides by UHPLC-HRMS.

## 2. Material and methods

### 2.1. Reagent and materials

Urea, tris(hydroxymethyl)aminomethane (Tris−HCl), tris(2-carboxyethyl)phosphine hydrochloride (TCEP), iodoacetamide (IAA), tetraethylammonium bicarbonate (TEAB), and NAP-10 columns (GE Healthcare) were purchased from Sigma-Aldrich (Bornem, Belgium). Trifluoroacetic acid (TFA) was obtained from Thermo Fisher Scientific (Waltham, Massachusetts, USA), and solid phase extraction cartridge (SPE) Pak C18 1 cc Vac Cartridge from Waters (Milford, Massachusetts, USA). Enzymatic digestion was performed with Trypsin Gold, Mass Spectrometry Grade from Promega (Madison, Wisconsin, USA), and Lysyl Endopeptidase, Mass Spectrometry Grade (Lys-C) from Wako Pure Chemical Industries, Ltd (Osaka, Japan). The water, acetonitrile, methanol absolute, 2-propanol (ULC/MS grade for all solvents), and formic acid were from Biosolve (Valkenswaard, the Netherlands). ESI-L Low Concentration Tuning Mix was purchased from Agilent Technologies (Santa Clara, California, USA)

### 2.2. Food product preparation

Processed egg matrices were produced in the Food Pilot unit of the Flanders Research Institute for Agriculture, Fisheries and Food (ILVO). The preparation of the egg matrices started with homogenized raw liquid whole egg (12.49% protein content) or homogenized pasteurized liquid whole egg (12.49% protein content). Neither egg product contained any additives. The eggs were purchased from chicken farms engaged in cage farming (Lodewijckx NV, Veerle, Belgium, egg code 3).

#### 2.2.1. Preparation of a raw egg matrix

Raw whole liquid egg was freeze-dried upon arrival (freeze dryer Epsilon 2–10 D LSC, Martin-Christ, Osterode am Harz, Germany), vacuum packed (vacuum packaging machine KN5, VC999, North Kansas City, Missouri, USA), and stored at 4 °C in the dark.

#### 2.2.2. Preparation of a heated egg matrix

Pasteurized whole liquid egg was freeze-dried upon arrival (freeze dryer Epsilon 2–10 D LSC), vacuum packed (vacuum packaging machine KN5, VC999), and stored at 4 °C in the dark. Egg pasteurization fulfils USDA requirements indicating that whole eggs must reach a minimum temperature of 60 °C for 3.5 min [16].

#### 2.2.3. Preparation of fatty rich environment egg matrix: type 1 = mayonnaise

A volume of 450 mL sunflower oil (Carrefour, Boulogne-Billancourt, France) was added slowly to 100 g raw liquid whole egg while being mixed (Bosch immersion blender MSM67170—750 W, Stuttgart, Germany) until a smooth mayonnaise was formed. The mayonnaise was divided into 50 g aliquots, vacuum packed (vacuum packaging machine KN5, VC999), and stored at 4 °C in the dark.

#### 2.2.4. Preparation of fatty rich environment egg matrix: type 2 = chocolate

First, a fine egg powder was produced by mixing 10 g freeze-dried raw egg powder into a fine powder (Kenwood Mini Chopper Mill Attachment AT320B [Hampshire, United Kingdom], pulsing speed 3 min) followed by the sieving of this powder (Retsch test

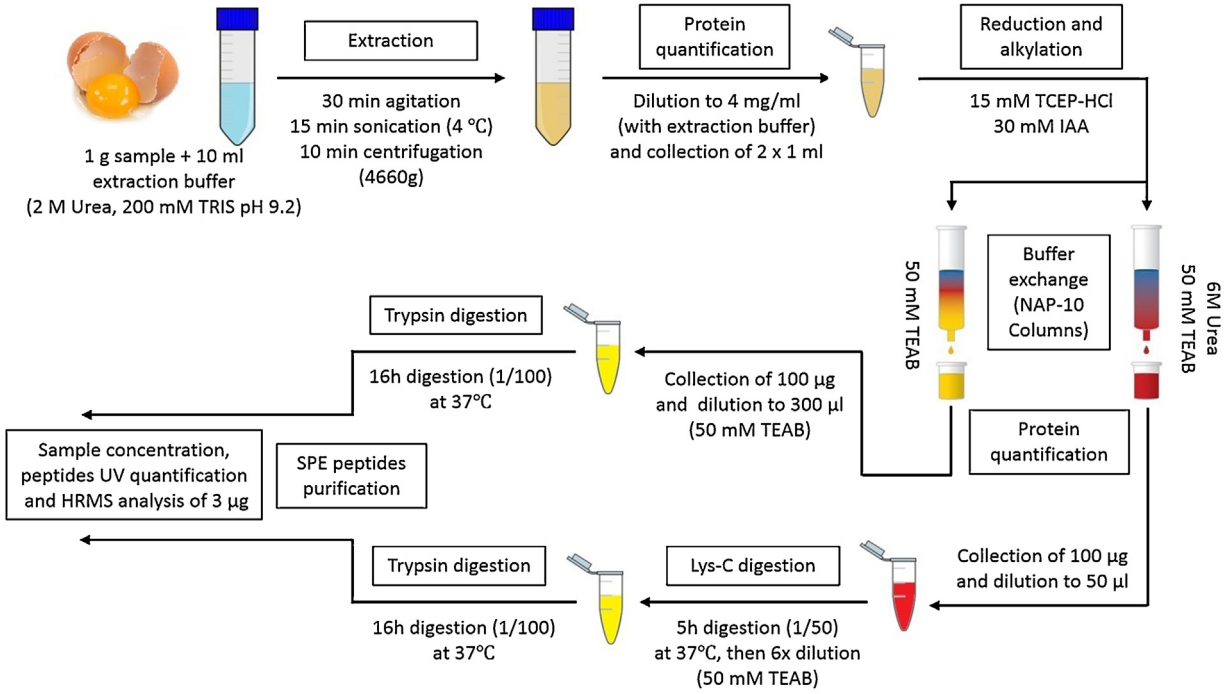



**Fig. 1.** Sample preparation flowchart. Proteins from a 1 g sample were extracted with 10 mL extraction buffer. The proteins were then reduced and alkylated. The buffer was exchanged with an appropriate buffer depending on the subsequent proteolytic digestion (trypsin alone or tandem Lys-C/trypsin). The resulting peptides were purified with the SPE before HRMS analysis was performed.

sieve [Haan, Germany], pore size 500 µm). Chocolate was then prepared. First, a "conche" was made from chocolate refiner flakes (90%) and cacao butter (10%). An equivalent of 450 g chocolate refiner flakes (Callebaut, Lebbeke, Belgium) and 50 g cacao butter (Callebaut) were warmed in a double boiler in separate bowls until a temperature of 40–45 °C was reached. The temperature was controlled with a food thermometer. The melted butter was then added slowly to the melted chocolate refiner flakes, and the mix was stirred for 3 min with a spoon. Next, 1.2% ammonium phosphatide (Palsgaard A/S; Juelsminde, Denmark) was added, and the mixture was stirred for 3 more minutes. Finally, a chocolate incurred with egg (45,454 ppm egg proteins in matrix) was produced by adding 50 g egg powder to the liquid chocolate and stirring again for 3 min while maintaining the temperature of the chocolate at 42 °C. The chocolate was subsequently poured into sterile chocolate molds, covered with plastic foil, and left to cool at room temperature. Once it had cooled, the chocolate was kept overnight at 4 °C to cool further, to set, and to solidify. The next day, the chocolate bars were rasped manually into chocolate flakes in 50 g aliquots to avoid the melting of the chocolate. Chocolate flakes were kept in plastic jars in 50 g aliquots and stored at 4 °C in the dark. A proper blank chocolate was prepared following the same procedure without adding the egg powder.

#### 2.2.5. Preparation of a fatty and low pH egg matrix: mayonnaise containing vinegar

A volume of 450 mL sunflower oil (Carrefour, Boulogne-Billancourt, France) was slowly added to 100 g raw liquid whole egg while being mixed (Bosch immersion blender MSM67170—750 W) until a smooth mayonnaise was formed. Next, 50 g vinegar (De Blauwe Hand, Puurs, Belgium) was mixed with the mayonnaise for 3 min (Bosch immersion blender MSM67170—750 W). The mayonnaise was divided into 50 g aliquots, vacuum packed (vacuum packaging machine KN5, VC999), and stored at 4 °C in the dark.

### 2.3. Sample preparation

The sample preparation protocol was based partly on the protocol of Planque et al. [10]. In short, proteins were extracted from food matrices, purified, and proteolytically digested. Two digestion protocols were compared for each processed food product. In the first protocol, the proteins were digested with trypsin alone. In the second, a tandem Lys-C/trypsin digestion was used. For each combination of digestion protocol and processed food product, three biological replicates were prepared and analyzed by UHPLC-HRMS (Fig. 1).

#### 2.3.1. Protein extraction

The proteins from a 1 g sample were extracted in 50 mL conical centrifuge tubes with 10 mL extraction buffer (2 M urea, 200 mM Tris—HCl; pH 9.2) by shaking at 20 °C for 30 min (Agitelec, Toulemonde, Mery-sur-Oise, France) followed by sonication (Elmasonic S 70, Elma Schmidbauer GmbH, Singen, Germany) for 15 min at 4 °C. The samples were centrifuged at 4660 g for 10 min at 4 °C. The middle liquid phase was then collected, and its protein concentration was measured using the Pierce 660 nm Assay (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

#### 2.3.2. Protein reduction, alkylation and purification

The samples were diluted to a 4 mg/mL protein concentration with extraction buffer. The proteins from a 1 mL aliquot were reduced and alkylated using TCEP (15 mM final concentration) and IAA (30 mM final concentration); they were then mixed (500 rpm) for 15 min at 37 °C in the dark. The samples were desalted and buffer-exchanged by gel filtration chromatography (NAP-10 columns, GE Healthcare, United Kingdom). The protease used for protein digestion in the next step determined the nature of the exchange buffer. A 50 mM TEAB buffer was employed for trypsin only digestion; 6 M urea and 50 mM TEAB were employed for the tandem Lys-C/trypsin digestion. The protein concentration of the

**Table 1**
Unique proteins and number of unique peptides for these proteins identified with HRMS analysis of egg samples (raw egg, heated egg, fatty rich environment, and low pH matrix). These unique peptides and proteins are identified in at least one sample using the following search parameters: "trypsin digestion", "allow up to three missed cleavages", "masses of monoisotopic peptide ions with a tolerance of 7 ppm", and "a fragment tolerance of 0.05 Da." Carbamidomethylation of the cysteines was set as a fixed modification. Oxidation of the methionine, conversion of the N-terminal glutamine into pyroglutamate, and deamidation of the asparagine and glutamine were allowed as variable modifications. The MS/MS-based protein and peptide identifications were validated using Scaffold 4.8. The peptide identification was accepted if a peptide could be established at >95% probability by the Peptide Prophet algorithms. The proteins were identified assuming a false discovery rate of 1% and accepted if they contained at least two identified peptides.

| Allergenic ingredient | Identified protein | UniProt entry identifier | Number of identified unique peptides | Coverage |
|---|---|---|---|---|
| Egg white | Alpha-1-acid glycoprotein | A7UEB0 | 6 | 27% |
| | Clusterin | Q9YGP0 | 6 | 10% |
| | Hep21 protein | Q8AV77 | 3 | 34% |
| | Lysozyme C (Gal d 4) | P00698 | 22 | 81% |
| | Mucin-5B | Q98UI9 | 24 | 12% |
| | Ovalbumin (Gal d 2) | P01012 | 36 | 59% |
| | Ovalbumin-related protein X | R9TNA6 | 14 | 34% |
| | Ovalbumin-related protein Y | P01014 | 24 | 34% |
| | OvoglobulinG2 | I0J170 | 6 | 10% |
| | Ovomucoid (Gal d 1) | P01005 | 5 | 41% |
| | Ovostatin | P20740 | 6 | 3% |
| | Ovotransferrin (Gal d 3) | P02789 | 86 | 71% |
| Egg yolk | Apolipoprotein B | F1NV02 | 166 | 27% |
| | Apovitellenin-1 | P02659 | 4 | 33% |
| | Ig lambda chain C region | P20763 | 4 | 20% |
| | Serum albumin (Gal d 5) | P19121 | 38 | 33% |
| | Vitellogenin-1 (Gal d 6) | P87498 | 81 | 33% |
| | Vitellogenin-2 | P02845 | 144 | 54% |

1.5 mL volume samples was again measured with the Pierce 660 nm Assay.

*2.3.3. Proteolytic digestion*

Two types of proteolytic digestion were used. For the trypsin-only digestion, the samples were diluted with 50 mM TEAB buffer to a 0.33 mg/mL protein concentration, and 300 μL of these samples was transferred to a 1.5 mL microcentrifuge tube. A 16 h digestion at 37 °C under 300 rpm agitation (Stuart® Microtiter plate shaker incubator SI505, Bibby Scientific, Stone, United Kingdom) was performed after 1 μg of trypsin (protein:trypsin ratio 1:100) was added. For the tandem Lys-C/trypsin digestion, the samples were diluted with 6 M urea and 50 mM TEAB buffer to 2 mg/mL protein concentration; 50 μL was transferred to a 1.5 mL microcentrifuge tube. A first 5 h digestion was performed at 37 °C under 300 rpm agitation by adding 2 μg of Lys-C. The samples were then diluted to a 300 μL final volume (to reduce the urea concentration to 1 M). A second 16 h digestion was performed at 37 °C under 300 rpm agitation by adding 1 μg of trypsin. For both digestion types, the reaction was stopped by the addition of 10 μL of 33% TFA.

To minimize potential carbamylation of proteins and peptides, urea solutions were freshly prepared. Extraction and digestion steps were also performed during the same day.

*2.3.4. SPE purification*

The peptides resulting from the digestion were purified using C18 SPE cartridges. The cartridges were first washed with 1 mL acetonitrile/water (50/50, v/v) and equilibrated with 1 mL wash solvent (acetonitrile/water/TFA, 2/97.9/0.1, v/v/v). The samples were diluted 10 times with water, loaded onto the cartridges, and washed with 1 mL wash solvent. The peptides were eluted with 2 mL acetonitrile/water (80/20, v/v). The samples were subsequently evaporated under a nitrogen flow at 40 °C. The pellets were dissolved in 500 μL of 0.1% formic acid and concentrated approximately 10 times with a vacuum concentrator (Heto, Denmark).

*2.4. Mass spectrometry*

The peptides were analyzed using an ESI-MS/MS maXis Impact UHR-TOF (Bruker, Billerica, Massachusetts, USA) coupled to a Dionex UltiMate 3000 Standard LC Systems (Thermo Fisher Scientific, Waltham, Massachusetts, USA). The peptide concentrations were measured using NanoDrop ultra-violet absorbance [17] (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Each injected sample volume was adapted such that each time, 3 μg peptides were analyzed. The peptides were separated by reverse-phase liquid chromatography on an Acclaim PepMap100 C18 column (3 μm, 100 A°, 1 mm × 15 cm nanoViper, Thermo Fisher Scientific, Waltham, Massachusetts, USA) using the Dionex UltiMate 3000 liquid chromatography system. The column temperature was set at 40 °C A and B mobile phases were water and acetonitrile, respectively, each containing 0.1% formic acid. The solvent gradient and the flow rate (FR) were set as follows: 0–3 min: FR 30 μL/min and 5% B; 3–40 min: FR 30 μL/min and 5%–40% B; 40–45 min: FR 40 μL/min and 40%–95% B; 45–50 min: FR 50 μL/min and 95% B; 50–50.1 m in. FR 30 μL/min and 95% to 5% B; and 50.1–60 min: FR 30 μL/min and 5% B. After each sample analysis, stringent washing was performed using water/acetonitrile/methanol/2-propanol (1/1/1/1) for 20 min to elute the residual components. Peptides of interest are mainly eluted within the 3%–40% acetonitrile gradient. But, to avoid contamination between the different samples, this strategy, including variable flow rate and stringent wash, was developed to elute the stickier peptides that remained on the liquid chromatography column.

The column effluent was connected to an electrospray ionization source. The MS and MS/MS data acquisition was carried out through 3 s cycles. Each cycle started with a 0.5 s survey scan during which MS spectra were acquired in a mass-to-charge ($m/z$) range of 150–2200. The most intense peptide ions containing 2–4 positive charges were selected for fragmentation during the remaining 2.5 s. The collision-induced dissociation (CID) energy was automatically set according to the $m/z$ ratio and the charge state of the precursor ion. Dynamic exclusion is used in the acquisition method: after a first spectrum acquisition, the ion is excluded during 2 min except if its intensity increases 3 times. The mass spectrometer and liquid chromatography systems were piloted by Compass Hystar 3.2 (Bruker, Billerica, Massachusetts, USA).

*2.5. Data analysis*

The peak lists were generated using DataAnalysis 4.2 (Bruker, Billerica, Massachusetts, USA). They were saved as MGF files. To

improve mass accuracy, the data were processed with a lock mass calibration [hexakis(1 h,1 h,4h-hexafluorobutyloxy)phosphazine, principal positive ion at $m/z$ = 1221.9906]. The MGF files were then treated with ProteinScape 3.1 (Bruker, Billerica, Massachusetts, USA) and Mascot 2.5 (Matrix Science, London, United Kingdom) as the search engine. The peak lists were searched against a homemade protein database downloaded from UniProt (uniprot.org). This database contains hen's proteome (*Gallus gallus*) and some possible contaminants (cacao, human keratin, trypsin and Lys-C) to avoid misidentification for a total of 77,682 entries (36,371 entries for hen, 40,959 for cacao, and 350 for human keratin). The following search parameters were used: "trypsin digestion", "allow up to three missed cleavages", "masses of monoisotopic peptide ions with a tolerance of 7 ppm", and "a fragment tolerance of 0.05 Da." Carbamidomethylation of the cysteines was set as a fixed modification. Oxidation of the methionine, conversion of the N-terminal glutamine into pyroglutamate, and deamidation of the asparagine and glutamine were allowed as variable modifications. Last, the peak lists were searched against the homemade database with an automatic decoy database search.

The MS/MS-based protein and peptide identifications were validated using Scaffold 4.8 (Proteome Software, Portland, Oregon, USA). The peptide identification was accepted if a peptide could be established at >95% probability by the Peptide Prophet algorithms [18,19]. The proteins were identified assuming a false discovery rate of 1% and accepted if they contained at least two identified peptides.

In parallel, the ProfileAnalysis (Bruker, Billerica, Massachusetts, USA) software was used to align the UHPLC-MS/MS raw data to compare the presence and the intensity of the peptides through the various samples. Using the "find molecular features" algorithm, the ions belonging to the same compound were combined into one feature (e.g., isotopes and charge states). The retention times of the obtained features were then aligned to rectify any potential retention shift that could appear during the run. Buckets were created, and a quantile normalization was applied. These buckets corresponded to the pairs of feature masses and the retention times. Comparable results were built by combining the list of peptides identified and validated by Scaffold and the buckets generated by ProfileAnalysis. The signal intensity linked to the identified peptides obtained this way was the basis for the results discussed below.

## 3. Results and discussion

### 3.1. Global strategy

Given the urgent need for a reference method for detecting and quantifying allergens in food [20], the objective of this study was to identify robust and specific egg peptide biomarkers that can be further used in a UHPLC-MS/MS method to quantify egg in processed food products. With the HRMS-based discovery proteomics approach used in this study, more than 600 egg peptides were identified in the samples. Unique identified proteins and the number of unique peptides for these proteins are listed in Table 1. The complete list of peptides associated with their intensity in each egg matrix can be found in Supplementary material 1. To ensure the sensitivity, specificity, robustness, and trueness of the future method, it was necessary that the selected analytical targets fulfil a series of criteria detailed in the following sections. The strategy was thus to filter, using these criteria, all of the identified peptides and to withhold the ideal biomarkers for the quantitative method.

### 3.2. Selection criterion 1: protein distribution in the egg

Eggs are composed of three main parts: egg white or albumen, yolk, and eggshell, which includes egg membrane. These three parts are used together by the food industry, but they can also be found individually in some recipes. To be able to detect egg in any situation, this issue needed to be considered in the selection of the biomarkers. Eggshell, generally considered a waste product, can be used as a food additive because of its high calcium content (>90%) [21]. Eggshell also contains about 6% protein. These proteins correspond to eggshell-specific proteins and to proteins found in egg white. The eggshell membrane is composed mainly of collagen and egg white proteins [22–24]. Because both the eggshell and the eggshell membrane contain egg white proteins, such as ovalbumin or lysozyme, and no specific allergen has been identified in these parts [25], the selected biomarkers in this study did not include shell-specific peptides. In contrast, both the egg white and the yolk, which can be used individually by the food industry, contain specific allergens [26]; thus, the peptide biomarkers had to be selected from these two parts.

### 3.3. Selection criterion 2: the amino acid composition of targeted peptides

For the quantitative UHPLC-MS/MS method, the peptide biomarker selection has to comply with specific rules to ensure method trueness. Indeed, the total egg protein quantification is directly linked to the concentration of the selected peptides through their respective protein quantities in the egg. The targeted peptides should therefore be found in a single form: a single mass. Thus, the peptides containing amino acids that are prone to modification needed to be omitted because such peptides could be present in non-modified or modified versions in not very predictable proportions. Based on a review of the literature, potential modifications were empirically identified using Mascot's "error tolerant search." The identified modifications are summarized in Supplementary material 2.

Cysteine modification is intentionally introduced by the sample preparation. Indeed, cysteine are reduced and alkylated to disrupt disulfide bridges and increase protein denaturation before the enzymatic digestion step. TCEP and IAA used for this purpose induce cysteine carbamidomethylation. This modification was set as a fixed modification for peptide identification in the data analysis. This was done because of the observed high modification rate and the need to limit the number of modifications set as a variable. This in turn increased the database search time and reduced its specificity [27]. Nevertheless, the decision was made to reject peptides containing cysteine to prevent the effects of possible incomplete carbamidomethylation, which could induce a bias in the additional quantitative method.

The other possible amino acid modifications that were identified were methionine oxidation, asparagine and glutamine deamidation and N-terminal glutamine-to-pyroglutamate conversion. Therefore, the peptides containing cysteine, methionine, or N-terminal glutamine were systematically rejected.

Deamidation was treated differently, given the high occurrence of asparagine and glutamine [28]. Deamidation is typically a spontaneous reaction converting asparagine (Asn) to aspartic acid or isoaspartic acid and glutamine (Gln) to glutamic acid. The reaction rate is influenced by temperature and pH and also by the nature of the amino acid on the carboxyl side adjacent to either the asparagine or glutamine [29–32]. The fastest reaction occurs when this carboxyl side adjacent amino acid is a glycine (Gly). For other amino acids, the reaction rates are at least 10 times slower [29,31]; thus, only the peptides containing Asn-Gly or Gln-Gly sequences were rejected. Last, post-translational modifications must also be taken into account. For example, serine phosphorylation occurs in egg proteins, such as ovalbumin. The peptides known to contain such modified amino acids were rejected. This information

has been documented in proteomics databases, such as UniProt (uniprot.org).

To assess the importance of considering these peptide modifications for the MS quantitative analysis, a systematic analysis of the signal intensity of the vitellogenin-2 peptides either unmodified (hereafter called native peptides) or modified by methionine oxidation, glutamine-to-pyroglutamate conversion, or glutamine and asparagine deamidation was performed. As was shown in the graph in Supplementary material 3, the proportion of native and modified peptides was hardly predictable. Variations were also observed among the different modifications, and food processing also seemed to have an effect. For example, methionine oxidation was observed in mayonnaise more than in unprocessed eggs. However, no general trend was identified. Peptide modifications were observed in the various egg proteins and for the different food processes considered in this study without predictable ratios between the native and modified peptides. Therefore, the peptides containing the methionine, cysteine, N-terminal glutamine, or Asn-Gly or Gln-Gly sequences were systemically rejected to ensure method trueness.

### 3.4. Selection criterion 3: protease(s) used and cleavage site composition

In the protocol of Planque et al., as in many other studies, the proteins were enzymatically digested with trypsin. To improve the digestion of the trypsin-resistant proteins, the use of Lys-C in combination with trypsin was tested. Trypsin is the most used protease in bottom-up mass spectrometry-based proteomics [33,34]. The reason is its high catalytic activity and specificity in peptide bonds hydrolysis at the C-terminal side of the arginine and lysine residues. Tryptic peptides are ideal for mass spectrometry analysis (length, hydrophobicity and ionization). Lys-C can be used in combination with trypsin, given their partly overlapping specificities (cleavage at the Cterminal side of lysine). One interesting advantage of Lys-C is that it works efficiently under highly denaturing conditions, thus also limiting the risk of missed cleavages that otherwise form a source of the underestimation of the peptide quantitation [35,36].

As was described in the material and methods section, the proteins were extracted and digested with trypsin alone or with a Lys-C/trypsin combination. Both digestion protocols were applied parallel to each sample. As is shown in Fig. 2, the improvement associated with Lys-C/trypsin digestion varied according to the specific protein being considered. The peptide MS signal intensity following the tandem Lys-C/trypsin digestion was significantly higher (compared to the trypsin alone) for most of the proteins from the egg white (ovalbumin [Fig. 2A], ovomucoid, ovotransferrin [Supplementary material 1]), and for the egg yolk proteins (Apolipoprotein B, Apovitellenin-1 [Supplementary material 1]). For some proteins, however, the digestion efficiency improvement was less obvious. No improvement was observed for lysozyme (Fig. 2B); however, for other proteins, such as vitellogenin-1 or vitellogenin-2 (Fig. 2C), an improvement was observed only for select peptides.

On the contrary, the food processing applied to the samples had a limited effect on the efficiency of the enzymatic digestion. The proteins that were resistant to trypsin digestion but were more efficiently digested with the tandem Lys-C/trypsin exhibited the same behavior independent of the food processing. The same observation was reported for the proteins without trypsin digestion resistance, such as lysozyme, or for the partly resistant proteins (Supplementary material 1).

According to the classic rules for trypsin and Lys-C proteolysis, trypsin cleaves at the carboxyl side of lysine and arginine whereas Lys-C cleaves only after lysine unless a proline immediately follows the cleavage site for both enzymes [37,38]. Additional factors are known to affect the digestion efficiency and the missed cleavage probability of a particular cleavage site. Because of the goal of using a quantitative UHPLC-MS/MS method, the cleavage of the selected peptide biomarkers had to be as complete as possible. Indeed, if a selected peptide is subjected to a higher possibility of missed cleavages, its absolute quantification will be biased, and the egg protein concentration will be underestimated.

Assuming optimized digestion conditions (pH, temperature, duration, etc.), the amino acid sequence surrounding the cleavage site has been found to affect peptide bond hydrolysis efficiency [37–41]. This information has to be taken into account for the peptide selection, and, whenever possible, peptides containing or surrounded by a resistant sequence have to be avoided. These sequences are described, using the nomenclature formulated by Schechter and Berger [42], as P4-P3-P2-P1-P1′-P2′-P3′, in which cleavage of the peptide bond occurs between P1 and P1′. As was previously mentioned, both proteases had a lower digestion rate for sequences with a proline at position P1′. Lysine (K) and arginine (R) also had a negative effect on cleavage efficiency when present at position P1′. In other words, R—R, R—K, K–R and K—K were prone to missed cleavage events. Finally, the negatively charged amino acid residues, aspartate (D) and glutamate (E) or phosphorylated serine (pS) and threonine (pT) had a negative influence when present close to the cleavage site. The amino acid residues affecting digestion efficiency are summarized in Table 2.

Whereas P, K, and R at the P1′ positions can be easily avoided without too much interference of the overall sensitivity of the method, totally avoiding D and E in the surrounding of the cleavage site is more cumbersome given their relatively high natural occurrence. This criterion was taken into account for peptide selection but not as an exclusive rule, otherwise too many peptides would be rejected. The peptides under consideration (i.e., those containing D and/or E in the area surrounding the cleavage site) were treated on a case-by-case basis, with the importance of the potential resulting missed cleavage evaluated. A peptide was rejected when the MS signal intensity of the associated peptide containing a missed cleavage exceeded 10% of the MS signal of the fully tryptic peptides.

### 3.5. Selection criterion 4: robustness of peptides to food processing

Eggs are usually consumed in the form of processed products. It is known that food processing conditions and matrix interactions affect the recovery and/or the further detectability of target molecules [43,44]. Such adverse conditions include heat treatment, the pH of the matrix, its fat content, and the presence of tannins. To ensure the robustness and the accuracy of the method, the effect of food processing was taken into account for the peptide biomarker selection. The matrices and the food processing applied to the samples containing egg were selected to represent the conditions found in everyday food products. Ideally, the peptides selected for quantitative UHPLC-MS/MS should not be affected.

The food processing effect was evaluated by comparing the MS signal intensity of the peptides under the various test conditions. Examples of vitellogenin-1 and ovotransferrin, egg yolk and egg white proteins, respectively, are shown in Fig. 3. In the case of vitellogenin-1 (Fig. 3A), some peptides, such as ATAVSLLEWQR or NVNFDGEILK, seemed resistant to the selected food processes. These peptides are thus ideal biomarkers for the quantitative UHPLC-MS/MS method. In contrast, other peptides are affected by the food processing like low pH for peptides ALLSEIR or YLLDLLPAAASHR. The situation is different for ovotransferrin (Fig. 3B). This protein seems to be unaffected by food processing investigated in this study. All of the peptides were detected at comparable intensities; thus, they could be potential biomarkers. In this case, the selected peptide biomarkers are those with the highest signal intensity to ensure an overall high sensitivity of the

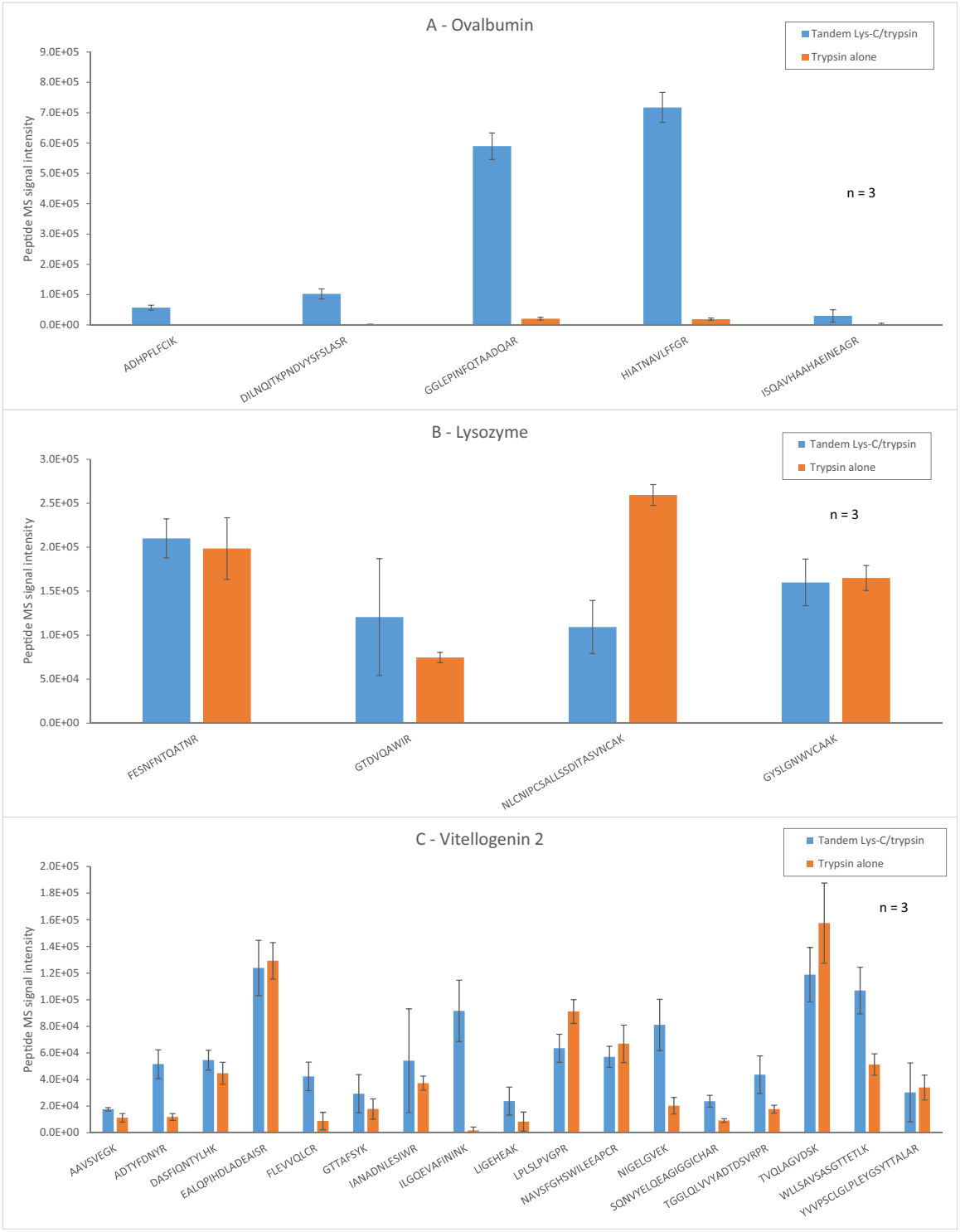



**Fig. 2.** Peptide MS signal intensity comparison between the two digestion protocols (tandem Lys-C/trypsin and trypsin alone) in unprocessed eggs for 3 proteins (A: ovalbumin; B: lysozyme; C: vitellogenin-2). Results are expressed as the means of the peptide MS signal intensity ± 1 SD for 3 independent replicates.

**Table 2**
Summary of amino acids favoring missed cleavages by trypsin and Lys-C. The relative position of these amino acids to the cleavage site is shown using Schechter and Berger's nomenclature [46].

| Relative position to cleavage site | | | | | | |
|---|---|---|---|---|---|---|
| P4 | P3 | P2 | P1 | P1′ | P2′ | P3′ |
| D E<br>pS pT | D E<br>pS pT | D E<br>pS pT | – | D E K P R<br>pS pT | D E<br>pS pT | D E<br>pS pT |

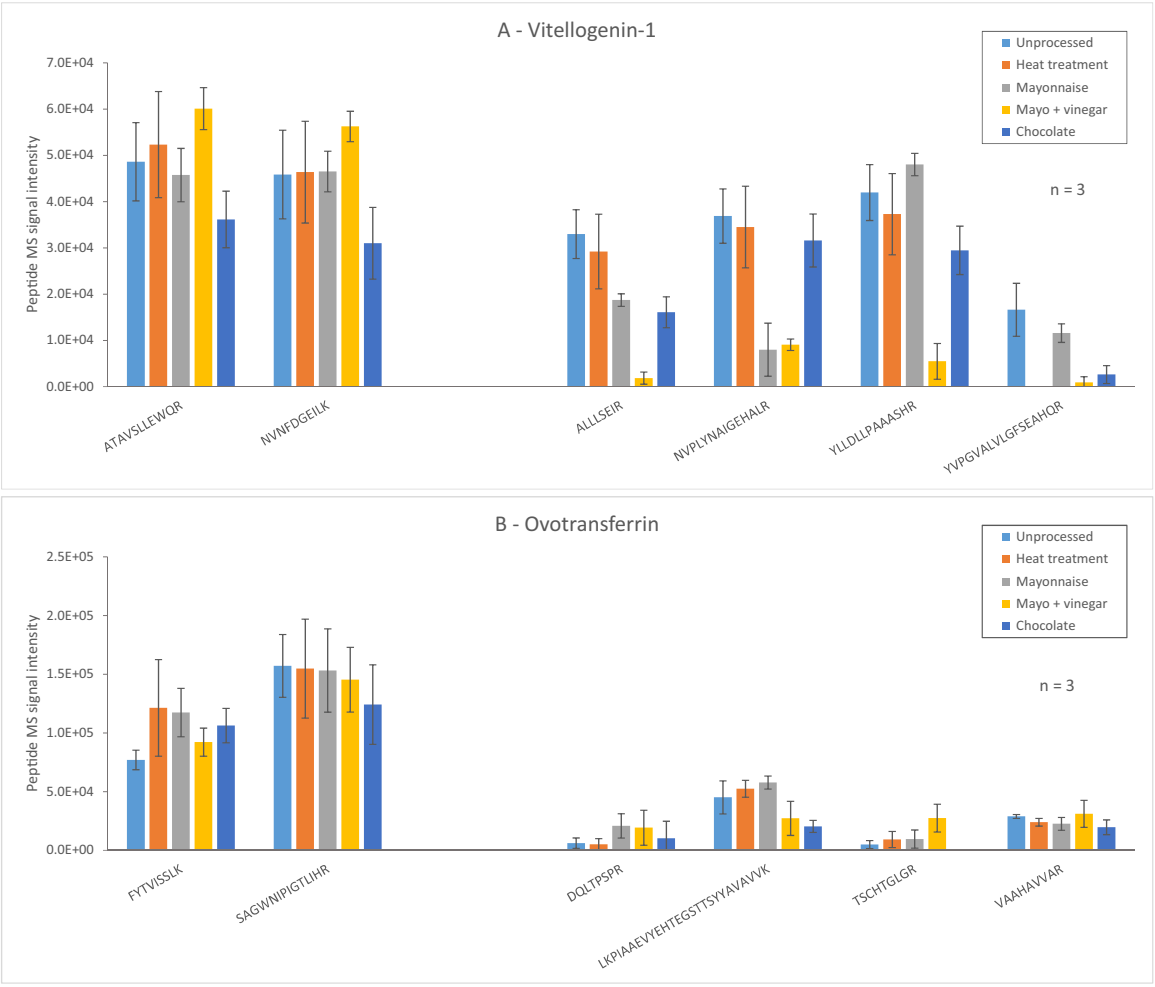



**Fig. 3.** Effect of food processing on a protein from egg yolk (A - Vitellogenin-1) and a protein from egg white (B - Ovotransferrin). Five matrices containing egg are compared. From left to right: no process, heated egg, fatty-rich environment (mayonnaise), acidic environment (mayo + vinegar), and the presence of tannins (chocolate). The comparison is based on the MS signal intensity. The ideal peptide biomarkers are located on the left of the figure. The results are expressed as the means of the peptide MS signal intensity ± 1 SD for 3 independent replicates.

UHPLC-MS/MS method. This should therefore allow the detection of low(er) amounts of egg proteins in processed food products.

Finally, the signal intensities of the peptides identified in the chocolate matrix were observed to be globally lower than those identified in the other matrices. This might be explained by matrix effects, such as the presence of tannins, which can interfere with sample preparation and HRMS analysis [9,45] or, simply, by the presence of cacao proteins. Indeed, proteins were extracted from the samples and digested, and 3 μg of the obtained peptides were analyzed by HRMS without consideration of their origin. Because the chocolate matrix was the only matrix containing endogenous proteins, the signals associated with the egg peptides were lower. This assumption was strengthened by the identification of peptides from cacao proteins, such as Vicilin-A and the 21 kDa seed protein.

### 3.6. Final selection of egg peptide biomarkers

Starting from more than 600 egg peptides identified by HRMS, the selection criteria discussed in the previous sections were systematically applied to identifying the ideal biomarkers. A list of 16 peptides from 6 proteins (3 from egg white and 3 from egg yolk) was obtained (Table 3). These peptides are potential biomarkers and will be used to develop a UHPLC-MS/MS method to quantify egg in processed food products.

### 3.7. Peptide specificity

Egg is one of the most important components in the human diet. Usually, it is hen eggs that are consumed; however, 7% of the entire egg production is from other poultry species [46], such as goose, duck, turkey, quail, and guinea fowl. Serologic and clinical cross-reactivity between hen eggs and other poultry eggs have been reported [47,48]. Thus, patients with hypersensitivity to hen eggs may also react allergically to other poultry eggs. The main goal of the method used in this study was to detect hen eggs, but there is also a need to detect eggs from other farm birds. Indeed, the European legislation regarding labelling clearly states that “egg” refers to eggs from all farm birds (Commission Notice of 13 July 2017 relating to the provision of information on substances or products causing allergies or intolerances as listed in Annex II to Regulation [EU] No 1169/2011 of the European Parliament and of the Council on the provision of food information to consumers). On the contrary, the selected peptides may not be identified in the proteins of other food ingredients otherwise, false positive results could be obtained. The sequences have to be specific to hen eggs, but they could possibly be extended to the eggs of other farm birds.

Given the empirical approach used in this study, it can be asserted that the selected peptide sequences are found in hen egg proteins. Indeed, the selection of egg peptides was based on the hen’s proteome. The specificity of these sequences to the eggs

**Table 3**
List of egg peptide biomarkers that fulfil the above-mentioned criteria. The peptides are linked to their respective proteins (with their positions in the protein sequences) and egg part (white or yolk). The protein accession numbers in the UniProt database have been placed in brackets.

| Allergenic ingredient | Protein | Selected egg peptide biomarkers |
|---|---|---|
| Egg White | Lysozyme (P00698) | $_{52}$FESNFNTQATNR$_{63}$ |
| | Ovalbumin (P01012) | $_{128}$GGLEPINFQTAADQAR$_{143}$ |
| | | $_{371}$HIATNAVLFFGR$_{382}$ |
| | | $_{141}$SAGWNIPIGTLIHR$_{154}$ |
| | Ovotransferrin (P02789) | $_{155}$GAIEWEGIESGSVEQAVAK$_{173}$ |
| | | $_{279}$VEDIWSFLSK$_{288}$ |
| | | $_{680}$FYTVISSLK$_{688}$ |
| Egg Yolk | | $_{257}$HLFLPSSYK$_{265}$ |
| | Apolipoprotein B (F1NV02) | $_{432}$ASFYGLSHAVTK$_{443}$ |
| | | $_{725}$ALFDYFGYSHDGK$_{737}$ |
| | | $_{840}$FALSGIVTPGAK$_{851}$ |
| | | $_{942}$TEEIPPLIENR$_{952}$ |
| | Apovitellenin 1 (P02659) | $_{75}$NFLINETAR$_{83}$ |
| | Vitellogenin 1 (P87498) | $_{1745}$TVIVEAPIHGLK$_{1756}$ |
| | | $_{1757}$NVNFDGEILK$_{1766}$ |
| | | $_{1874}$ATAVSLLEWQR$_{1884}$ |

**Table 4**
Poultry inter-species homology of selected egg peptide biomarkers using BLAST analysis with the UniProtKB database from UniProt and the non-redundant protein sequence database from NCBI. "X" indicates that the peptide was sequenced in the species, and X* that the sequence was predicted in the species.

| | Hen (*Gallus gallus*) | Goose (*Anser cygnoides domesticus*) | Duck (*Anas platyrhynchos*) | Turkey (*Meleagris gallopavo*) | Quail (*Coturnix japonica*) | Guinea fowl (*Numida meleagris*) |
|---|---|---|---|---|---|---|
| GGLEPINFQTAADQAR | X | | | | | |
| HIATNAVLFFGR | X | | | | | |
| FYTVISSLK | X | | | | | |
| GAIEWEGIESGSVEQAVAK | X | | | | | |
| SAGWNIPIGTLIHR | X | X* | X | | | X |
| VEDIWSFLSK | X | | | | | |
| FESNFNTQATNR | X | | | | X | |
| ATAVSLLEWQR | X | X | X | X* | X* | |
| NVNFDGEILK | X | | | | | X |
| TVIVEAPIHGLK | X | | | | | |
| ALFDYFGYSHDGK | X | | | X | X* | |
| ASFYGLSHAVTK | X | | | X | X* | X |
| FALSGIVTPGAK | X | | | X* | | X |
| HLFLPSSYK | X | X* | X | X* | X* | X |
| TEEIPPLIENR | X | X* | X | X* | X* | |
| NFLINETAR | X | | | | | |

of farmed birds and the identification of inter-poultry species homologies were assessed using Basic Local Alignment Search Tool (BLAST) analyses using the UniProt and NCBI protein databases. The results obtained for the poultry species are shown in Table 4. The GGLEPINFQTAADQAR peptide from the ovalbumin was specific for hen egg whereas the ATAVSLLEWQR peptide from vitellogenin 1 was found in several poultry species. This information allows for the development of methods specific to hen eggs or methods for differentiating the eggs of poultry species. Concerning the specificity to poultry egg, none of the selected peptides was found in the sequence of proteins of other food ingredients. For example, Peptide HLFLPSSYK has been identified in the protein of turtle, alligator, and snake eggs; however, because these eggs are not usually consumed, this has not been a problem. Thus, the procedure in this study ensures specificity, with regard to the databases considered. Indeed, all of the proteins in potential food ingredients have not yet been sequenced.

### 3.8. Comparison of selected peptide biomarkers with literature

The 16 egg peptides selected with our empirical approach were finally compared with egg peptide biomarkers described in the literature. In the table placed in Supplementary material 4, our potential biomarkers are highlighted in bold. The matrices used by authors are also referenced in this table.

A total of 54 egg peptides were identified in the literature, 44 from egg white proteins and 10 from egg yolk proteins. This tendency for peptides from egg white proteins is partially due to methods detecting egg white used for some winemaking techniques. But, as explained above, peptide biomarkers have to be selected from both egg white and egg yolk proteins to be able to detect egg in any situation.

Out of these 54 peptides from the literature, 44 were identified with our HRMS based method. Many of them would have been rejected by our selection criteria. For example, ovalbumin peptide LTEWTSSNVMEER is considered as egg biomarker in 11 publications. But, since it contains a potentially modified methionine, it was rejected for our future quantitative method. In contrast, 9 peptides of our final biomarkers selection aren't considered in other publications. The majority of them (8/9) are coming from egg yolk proteins.

## 4. Conclusions

Food allergy is a growing health problem; thus, there is an urgent need for accurate detection and quantification methods of the allergenic ingredients in processed food. Mass spectrometry-based methods are promising. They can overcome the limitations of classic detection methods, such as ELISA or PCR. The proteins of allergenic ingredients are detected and quantified through their

constitutive peptides, and the selection of these peptides is the first, and crucial, step in method development.

This study presented an empirical approach for identifying potential peptide biomarkers to detect and to quantify egg in processed food. The approach was based on the analysis of processed egg matrices (raw egg, heated egg, fatty rich environment, and low pH matrix) with HRMS. The identified peptides were filtered using a series of criteria to ensure method specificity, trueness, and robustness. Thus, peptides must be specific for the allergenic ingredient, be robust to food processing, have originated from abundant proteins, be size-compatible with QqQ, and not be prone to amino acid modifications or missed cleavages by the protease used. The use of a tandem Lys-C/trypsin enzymatic digestion improved the digestion of some of the proteins as compared to digestion with trypsin alone.

Considering these criteria, a list of 16 robust potential peptide biomarkers was obtained, allowing for the detection of the egg yolk and egg white proteins for eggs from hens and other farm birds (goose, duck, turkey, quail, and guinea fowl). In future research, these peptides will be used as targets in a quantitative UHPLC-MS/MS method for detecting food allergens.

## Acknowledgments

The research that yielded these results was funded by the Belgian Federal Public Service of Health, Food Chain Safety and Environment through the contract RT 15/10 ALLERSENS.

## Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.chroma.2018.11.036.

## References

[1] B.I. Nwaru, L. Hickstein, S.S. Panesar, A. Muraro, T. Werfel, V. Cardona, A.E.J. Dubois, S. Halken, K. Hoffmann-Sommergruber, L.K. Poulsen, G. Roberts, R. Van Ree, B.J. Vlieg-Boerstra, A. Sheikh, The epidemiology of food allergy in Europe: a systematic review and meta-analysis, Allergy Eur. J. Allergy Clin. Immunol. 69 (2014) 62–75, http://dx.doi.org/10.1111/all.12305.

[2] A. Muraro, T. Werfel, K. Hoffmann-Sommergruber, G. Roberts, K. Beyer, C. Bindslev-Jensen, V. Cardona, A. Dubois, G. duToit, P. Eigenmann, M. Fernandez Rivas, S. Halken, L. Hickstein, A. Høst, E. Knol, G. Lack, M.J. Marchisotto, B. Niggemann, B.I. Nwaru, N.G. Papadopoulos, L.K. Poulsen, A.F. Santos, I. Skypala, A. Schoepfer, R. Van Ree, C. Venter, M. Worm, B. Vlieg-Boerstra, S. Panesar, D. de Silva, K. Soares-Weiser, A. Sheikh, B.K. Ballmer-Weber, C. Nilsson, N.W. de Jong, C.A. Akdis, EAACI Food Allergy and Anaphylaxis Guidelines Group, EAACI food allergy and anaphylaxis guidelines: diagnosis and management of food allergy, Allergy 69 (2014) 1008–1025, http://dx.doi.org/10.1111/all.12429.

[3] S.H. Sicherer, H.A. Sampson, Food allergy: epidemiology, pathogenesis, diagnosis, and treatment, J. Allergy Clin. Immunol. 133 (2014), http://dx.doi.org/10.1016/j.jaci.2013.11.020.

[4] B. Popping, C. Diaz-Amigo, K. Hoenicke, Molecular Biological and Immunological Techniques and Applications for Food Chemists, 2010, http://dx.doi.org/10.1002/9780470637685.

[5] A.J. Van Hengel, Food allergen detection methods and the challenge to protect food-allergic consumers, Anal. Bioanal. Chem. 389 (2007) 111–118, http://dx.doi.org/10.1007/s00216-007-1353-5.

[6] Nomenclature for automated and mechanised analysis (Recommendations 1989), Pure Appl. Chem. 61 (1989) 1657, http://dx.doi.org/10.1351/pac198961091657.

[7] D.C. Liebler, L.J. Zimmerman, Targeted quantitation of proteins by mass spectrometry, Biochemistry 52 (2013) 3797–3806, http://dx.doi.org/10.1021/bi400110b.

[8] J. Heick, M. Fischer, B. Pöpping, First screening method for the simultaneous detection of seven allergens by liquid chromatography mass spectrometry, J. Chromatogr. A 1218 (2011) 938–943, http://dx.doi.org/10.1016/j.chroma.2010.12.067.

[9] M. Planque, T. Arnould, M. Dieu, P. Delahaut, P. Renard, N. Gillard, Liquid chromatography coupled to tandem mass spectrometry for detecting ten allergens in complex and incurred foodstuffs, J. Chromatogr. A 1530 (2017) 138–151, http://dx.doi.org/10.1016/j.chroma.2017.11.039.

[10] M. Planque, T. Arnould, M. Dieu, P. Delahaut, P. Renard, N. Gillard, Advances in ultra-high performance liquid chromatography coupled to tandem mass spectrometry for sensitive detection of several food allergens in complex and processed foodstuffs, J. Chromatogr. A 1464 (2016) 115–123, http://dx.doi.org/10.1016/j.chroma.2016.08.033.

[11] A.I. Sancho, E.N.C. Mills, Proteomic approaches for qualitative and quantitative characterisation of food allergens, Regul. Toxicol. Pharmacol. 58 (2010) S42–S46, http://dx.doi.org/10.1016/j.yrtph.2010.08.026.

[12] M. Rauh, LC-MS/MS for protein and peptide quantification in clinical chemistry, J. Chromatogr. B: Anal. Technol. Biomed. Life Sci. 883–884 (2012) 59–67, http://dx.doi.org/10.1016/j.jchromb.2011.09.030.

[13] N.R. Kitteringham, R.E. Jenkins, C.S. Lane, V.L. Elliott, B.K. Park, Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics, J. Chromatogr. B: Anal. Technol. Biomed. Life Sci. 877 (2009) 1229–1239, http://dx.doi.org/10.1016/j.jchromb.2008.11.013.

[14] S.H. Sicherer, H.A. Sampson, 9. Food allergy, J. Allergy Clin. Immunol. 117 (2006), http://dx.doi.org/10.1016/j.jaci.2005.05.048.

[15] M. Eggesbø, G. Botten, R. Halvorsen, P. Magnus, The prevalence of allergy to egg: a population-based study in young children, Allergy Eur. J. Allergy Clin. Immunol. 56 (2001) 403–411, http://dx.doi.org/10.1034/j.1398-9995.2001.056005403.x.

[16] USDA, Code of Federal Regulations, Section 590.570 - Pasteurization of liquid eggs, 2012 https://www.gpo.gov/fdsys/pkg/CFR-2012-title9-vol2/xml/CFR-2012-title9-vol2-sec590-570.xml.

[17] C.M. Stoscheck, Quantitation of protein, in: M.P.B.T.-M, E. Deutscher (Eds.), Guid. to Protein Purif., Academic Press, 1990, pp. 50–68, http://dx.doi.org/10.1016/0076-6879(90)82008-P.

[18] A.I. Nesvizhskii, A. Keller, E. Kolker, R. Aebersold, A statistical model for identifying proteins by tandem mass spectrometry, Anal. Chem. 75 (2003) 4646–4658, http://dx.doi.org/10.1021/ac0341261.

[19] A. Keller, A.I. Nesvizhskii, E. Kolker, R. Aebersold, Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search, Anal. Chem. 74 (2002) 5383–5392, http://dx.doi.org/10.1021/ac025747h.

[20] M.J. Walker, D.T. Burns, C.T. Elliott, M.H. Gowland, E.N.C. Mills, Is food allergen analysis flawed? Health and supply chain risks and a proposed framework to address urgent analytical needs, Analyst 141 (2016) 24–35, http://dx.doi.org/10.1039/c5an01457c.

[21] W. Daengprok, W. Garnjanagoonchorn, Y. Mine, Fermented pork sausage fortified with commercial or hen eggshell calcium lactate, Meat Sci. 62 (2002) 199–204, http://dx.doi.org/10.1016/S0309-1740(01)00247-9.

[22] E.C.Y. Li-Chan, H.-O. Kim, Structure and chemical compositions of eggs, in: Egg Biosci. Biotechnol., John Wiley & Sons, Inc., 2008, pp. 1–95, http://dx.doi.org/10.1002/9780470181249.ch1.

[23] M.T. Hincke, J. Gautron, M. Panheleux, J. Garcia-Ruiz, M.D. McKee, Y. Nys, Identification and localization of lysozyme as a component of eggshell membranes and eggshell matrix, Matrix Biol. 19 (2000) 443–453, http://dx.doi.org/10.1016/S0945-053X(00)00095-0.

[24] M. Wong, M.J.C. Hendrix, K. von der Mark, C. Little, R. Stern, Collagen in the egg shell membranes of the hen, Dev. Biol. 104 (1984) 28–36, http://dx.doi.org/10.1016/0012-1606(84)90033-2.

[25] M. Ebisawa, H. Tachimoto, K. Ikematsu, C. Sugizaki, Y. Masuda, M. Kimura, Allergenic activity of non-calcinated egg shell calcium, Arerugi 54 (2005) 471–477.

[26] B.J. Walsh, D. Barnett, R.W. Burley, C. Elliott, D.J. Hill, M.E.H. Howden, New allergens from hen's egg white and egg yolk, Int. Arch. Allergy Immunol. 87 (1988) 81–86, http://dx.doi.org/10.1159/000234653.

[27] J.S. Cottrell, Protein identification using MS/MS data, J. Proteom. 74 (2011) 1842–1851, http://dx.doi.org/10.1016/j.jprot.2011.05.014.

[28] T.P. Coultate, R.S. of C. (Great Britain), Food: The Chemistry of its Components, Royal Society of Chemistry, 2009 https://books.google.be/books?id=KF2A8Cz7B-cC.

[29] N.E. Robinson, A.B. Robinson, Molecular clocks, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 944–949, http://dx.doi.org/10.1073/pnas.98.3.944.

[30] B. Li, R.L. Schowen, E.M. Topp, R.T. Borchardt, Effect of N-1 and N-2 residues on peptide deamidation rate in solution and solid state, AAPS J. 8 (2006) E166–E173, http://dx.doi.org/10.1208/aapsj080120.

[31] A.A. Kosky, V. Dharmavaram, G. Ratnaswamy, M.C. Manning, Multivariate analysis of the sequence dependence of asparagine deamidation rates in peptides, Pharm. Res. 26 (2009) 2417–2428, http://dx.doi.org/10.1007/s11095-009-9953-8.

[32] N.E. Robinson, A.B. Robinson, Prediction of primary structure deamidation rates of asparaginyl and glutaminyl peptides through steric and catalytic effects, J. Pept. Res. 63 (2004) 437–448, http://dx.doi.org/10.1111/j.1399-3011.2004.00148.x.

[33] J. Bunkenborg, G. Espadas, H. Molina, Cutting edge proteomics: benchmarking of six commercial trypsins, J. Proteome Res. 12 (2013) 3631–3641, http://dx.doi.org/10.1021/pr4001465.

[34] L. Tsiatsiani, A.J.R. Heck, Proteomics beyond trypsin, FEBS J. 282 (2015) 2612–2626, http://dx.doi.org/10.1111/febs.13287.

[35] P. Giansanti, L. Tsiatsiani, T.Y. ew Low, A.J.R. Heck, Six alternative proteases for mass spectrometry-based proteomics beyond trypsin, Nat. Protoc. 11 (2016) 993–1006, http://dx.doi.org/10.1038/nprot.2016.057.

[36] T. Glatter, C. Ludwig, E. Ahrné, R. Aebersold, A.J.R. Heck, A. Schmidt, Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin

proteolysis over trypsin digestion, J. Proteome Res. 11 (2012) 5145–5156, http://dx.doi.org/10.1021/pr300273g.
[37] J. Kay, Specificity of proteolysis, FEBS Lett. 331 (1993) 201, http://dx.doi.org/10.1016/0014-5793(93)80331-N.
[38] R. Raijmakers, P. Neerincx, S. Mohammed, A.J.R. Heck, Cleavage specificities of the brother and sister proteases Lys-C and Lys-N, Chem. Commun. 46 (2010) 8827, http://dx.doi.org/10.1039/c0cc02523b.
[39] J.A. Siepen, E.J. Keevil, D. Knight, S.J. Hubbard, Prediction of missed cleavage sites in tryptic peptides aids protein identification in proteomics, J. Proteome Res. 6 (2007) 399–408, http://dx.doi.org/10.1021/pr060507u.
[40] J. Rodriguez, Does trypsin cute before proline? J. Proteome Res. 7 (2008) 300–305, http://dx.doi.org/10.1021/pr0705035.
[41] P.D. Gershon, Cleaved and missed sites for trypsin, Lys-C, and Lys-N can be predicted with high confidence on the basis of sequence context, J. Proteome Res. 13 (2014) 702–709, http://dx.doi.org/10.1021/pr400802z.
[42] E. Harper, A. Berger, On the size of the active site in proteases: pronase, Biochem. Biophys. Res. Commun. 46 (1972) 1956–1960, http://dx.doi.org/10.1016/0006-291X(72)90076-9.
[43] S.E. Khuda, L.S. Jackson, T.J. Fu, K.M. Williams, Effects of processing on the recovery of food allergens from a model dark chocolate matrix, Food Chem. 168 (2015) 580–587, http://dx.doi.org/10.1016/j.foodchem.2014.07.084.
[44] C.H. Parker, S.E. Khuda, M. Pereira, M.M. Ross, T.J. Fu, X. Fan, Y. Wu, K.M. Williams, J. DeVries, B. Pulvermacher, B. Bedford, X. Zhang, L.S. Jackson, Multi-allergen quantitation and the impact of thermal treatment in industry-processed baked goods by ELISA and liquid chromatography–tandem mass spectrometry, J. Agric. Food Chem. 63 (2015) 10669–10680, http://dx.doi.org/10.1021/acs.jafc.5b04287.
[45] K.J. Shefcheck, J.H. Callahan, S.M. Musser, Confirmation of peanut protein using peptide markers in dark chocolate using liquid chromatography–tandem mass spectrometry (LC-MS/MS), J. Agric. Food Chem. 54 (2006) 7953–7959, http://dx.doi.org/10.1021/jf060714e.
[46] E.S. Polat, O.B. Citil, M. Garip, Fatty acid composition of yolk of nine poultry species kept in their natural environment, Anim. Sci. Pap. Rep. 31 (2013) 363–368.
[47] T. Langeland, A clinical and immunological study of allergy to hen's egg white: VI. Occurrence of proteins cross-reacting with allergens in hen's egg white as studied in egg white from Turkey, duck, goose, seagull, and in hen egg yolk, and hen and chicken sera and fles, Allergy 38 (1983) 399–412, http://dx.doi.org/10.1111/j.1398-9995.1983.tb05083.x.
[48] C. Alessandri, M. Calvani, L. Rosengart, C. Madella, Anaphylaxis to quail egg, Allergy Eur. J. Allergy Clin. Immunol. 60 (2005) 128–129, http://dx.doi.org/10.1111/j.1398-9995.2004.00449.x.

### 4.2. "High-resolution mass spectrometry-based selection of peanut peptide biomarkers considering food processing and market type variation" (Gavage *et al*, 2020b)

The strategy to identify potential peptide biomarkers was then applied to peanut. Peanuts are among the most common food ingredients that induce severe allergic reactions and the most frequent cause of food-induced anaphylaxis, the most severe form of allergic reaction, that is life-threatening (Husain & Schwartz, 2012).

Peanut are commonly classified into four market types (Runner, Spanish, Virginia and Valencia) corresponding to different botanical varieties. The relative abundance of peanut allergens has been shown to vary depending on the growing conditions (Walczyk *et al*, 2013), as well as the specific cultivar (Wu *et al*, 2016). Moreover, some of these allergens do not correspond to unique protein sequences but include several protein isoforms (Palladino, 2018). Multiple protein isoforms and origin variation issues were considered by analysing peanuts from two geographical regions corresponding to two market types (Spanish and Virginia).

The developed and optimized sample preparation protocol was applied to processed peanut matrices: ground peanuts as unprocessed matrix, oven roasted ground peanuts as heated matrix, fermented peanut milk as low pH matrix, caramelised ground peanuts as a matrix subjected to Maillard reactions and chocolate incurred with peanuts as fat-rich and complex matrix. Resulting peptides were analysed by HPLC-HRMS. Each peanut matrix was produced and analysed in duplicate, one for each peanut market type.

Identified peptides were filtered using a set of selection criteria to ensure method specificity, sensitivity, and robustness. Ideal peptide biomarkers must be specific to peanuts, belong to abundant proteins, be size-compatible with triple quadrupole mass spectrometry analysis, and be robust to food processing but not prone to missed cleavages by the protease(s) used or to amino acid modifications.

This approach led to the identification of 16 potential peanut peptide biomarkers from two peanut allergens Ara h 1 and Ara h 3. No significant differences were observed between the two market types concerning the amount of Ara h 1 and its three isoforms. On the contrary, the abundance of some of the 12 Ara h 3 isoforms was found to vary according to the peanut market types. Our selection included peptides covering all protein isoforms of Ara h 1 and Ara h 3.

The 16 identified potential peptide biomarkers were considered for the development of the quantitative UHPLC-MS/MS analysis method.

The research article was published in *Food Chemistry* Journal.

Food Chemistry 304 (2020) 125428

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

# High-resolution mass spectrometry-based selection of peanut peptide biomarkers considering food processing and market type variation

Maxime Gavage[a,b], Kaatje Van Vlierberghe[c], Christof Van Poucke[c], Marc De Loose[c], Kris Gevaert[d,e], Marc Dieu[b,f], Patsy Renard[b,f], Thierry Arnould[b], Nathalie Gillard[a,*]

[a] CER Groupe, Rue du Point du Jour 8, 6900 Marloie, Belgium
[b] Laboratory of Biochemistry and Cell Biology (URBC)-Namur Research Institute for Life Sciences (NARILIS), University of Namur, 61, Rue de Bruxelles, 5000 Namur, Belgium
[c] Flanders Research Institute for Agriculture, Fisheries and Food (ILVO), Brusselsesteenweg 370, 9090 Melle, Belgium
[d] VIB Center for Medical Biotechnology, Albert Baertsoenkaai 3, B9000 Ghent, Belgium
[e] Department of Biomolecular Medicine, Ghent University, Albert Baertsoenkaai 3, B9000 Ghent, Belgium
[f] MaSUN, Mass Spectrometry Facility, University of Namur, 61, Rue de Bruxelles, 5000 Namur, Belgium

ARTICLE INFO

*Keywords:*
High-resolution mass spectrometry
Peptide biomarker selection
Processed food products
Multiple allergen isoforms
Peanut origin

ABSTRACT

To protect allergic patients and guarantee correct food labeling, robust, specific and sensitive detection methods are urgently needed. Mass spectrometry (MS)-based methods could overcome the limitations of current detection techniques. The first step in the development of an MS-based method is the identification of biomarkers, which are, in the case of food allergens, peptides. Here, we implemented a strategy to identify the most salient peptide biomarkers in peanuts. Processed peanut matrices were prepared and analyzed using an untargeted approach via high-resolution MS. More than 300 identified peptides were further filtered using selection criteria to strengthen the analytical performance of a future, routine quantitative method. The resulting 16 peptides are robust to food processing, specific to peanuts, and satisfy sequence-based criteria. The aspect of multiple protein isoforms is also considered in the selection tree, an aspect that is essential for a quantitative method's robustness but seldom, if ever, considered.

## 1. Introduction

Food allergies are a global health problem. Several recent studies have indicated that the prevalence of food allergies has increased in the last few decades, and they primarily affect populations in industrialized countries. Children seem to be affected more often than adults as it is estimated that immunoglobulin-E (IgE)-associated food allergies affect 3–8% of children and 1–3% of adults (De Silva et al., 2014; Longo, Berti, Burks, Krauss, & Barbi, 2013; Sicherer & Sampson, 2014). Peanuts (*Arachis hypogaea*) are among the most common foods that induce severe allergic reactions and the most frequent cause of food-induced anaphylaxis (Husain & Schwartz, 2012). Sixteen protein allergens have been identified in peanuts and termed Ara h 1–17 according to the systematic allergen nomenclature approved by the World Health Organization and the International Union of Immunological Societies' Allergen Nomenclature Sub-committee; Ara h 3 and Ara h 4 were *a posteriori* considered to be the same allergen and are both referred to as Ara h 3 (Pomés et al., 2018). The most common peanut allergens recognized by the serum IgE of more than 50% of allergic patients are Ara h 1–3 and Ara h 6. These four allergens are also the most abundant proteins in peanuts and are all seed storage proteins (Schmidt et al., 2009). Some of these 16 Ara h allergens do not correspond to unique protein sequences but include several protein isoforms. These multiple isoforms are expressed from different genes or are the result of different types of post-translational proteolytic processing of the same protein (Palladino, 2018).

Peanut species are classified into two subspecies, which are further classified into six botanical varieties based on morphology and growth habits (Moretzsohn et al., 2004). However, they are more commonly classified into four market types, including Runner, Spanish, Virginia and Valencia. The relative abundance of peanut allergens has been shown to vary depending on the growing conditions (Walczyk et al., 2013), as well as the specific cultivar (Wu et al., 2016). One study quantified the major allergen content of peanuts from different market types (Koppelman et al., 2016). In this report, the proportion of different Ara h allergens in peanuts was measured, and it was found that

* Corresponding author.
*E-mail address:* n.gillard@cergroupe.be (N. Gillard).

https://doi.org/10.1016/j.foodchem.2019.125428
Received 3 June 2019; Received in revised form 14 August 2019; Accepted 24 August 2019
Available online 26 August 2019

Ara h 1, 2, 3 and Ara h 6 account for 17.1 ± 3.4%, 6.2 ± 1.3%, 70.6 ± 8.6%, and 5.8 ± 1.8% of the total of extracted protein, respectively. The extraction yield of the 20 peanut samples analyzed in that study was 74.7 ± 8.5%.

Strict peanut avoidance remains the best option for allergic patients, despite progress being made in desensitization therapy (Bégin et al., 2014). This avoidance strategy requires rigorous and accurate food labeling and, consequently, sensitive and reliable detection methods to exclude/identify unintentional contamination. Commercially available detection methods are mainly based on immunological methods targeting specific proteins, such as enzyme-linked immunosorbent assays (ELISA). Other methods rely on polymerase chain reactions (PCRs) targeting allergen-specific DNA molecules. Despite their high sensitivity, both methods have their limitations. The antibodies used in immunological methods recognize epitopes. One drawback is the possible recognition by the antibodies of similar structures in non-allergenic ingredients, thus generating false positive results. In contrast, protein denaturation upon food processing can modify the conformation of the targeted proteins or the targeted epitopes, leading to false negative results (Van Hengel, 2007). Due to technical advances, MS-based methods have become the methods of choice for allergen detection (Monaci, Pilolli, De Angelis, & Mamone, 2015; Popping & Godefroy, 2011). Multi-allergen detection and quantification methods, including those for peanuts, were recently proposed (Boo, Parker, & Jackson, 2018; Henrottin et al., 2019; Planque et al., 2019). Studies on allergen detection by MS-based methods report detection limits in the same range (0.1–5 mg of allergenic ingredient protein per kg of matrix) as ELISAs (Croote & Quake, 2016). Lastly, the high specificity of MS-based methods is ensured by different technical elements including BLAST query of the selected peptides to public databases, chromatographic retention time or the analysis of multiple transitions.

The first step in the development of an MS-based method is the identification and further selection of possible peptides and, thus, proteins specific to the studied species (i.e., biomarkers). This selection is particularly important when developing a quantitative method since quantifying the allergenic ingredient will be directly linked to quantifying the selected biomarkers. Therefore, the method's overall accuracy, robustness, and precision rely on this selection. In a previous study (Gavage et al., 2018), we developed a strategy for identifying potential peptide biomarkers in processed food products. Our strategy, based on an untargeted approach using high-resolution mass spectrometry (HRMS), was applied to processed egg-containing matrices and allowed us to identify 16 potential egg peptide biomarkers.

The aim of the present study was to apply the same approach to identifying potential biomarkers of peanut peptides. Modifications needed to be implemented to adjust our approach to address peanuts and consider relevant food matrices that best represent processed foods that potentially contain peanuts. To include potential peanut origin variability, peanuts from two distinct geographical regions corresponding to two different market types were analyzed. We identified more than 300 peptides, which were further filtered using a series of selection criteria to ensure the adequate performance of a future, routine ultra-high-performance liquid chromatography (UHPLC)-MS/MS analytical method. These criteria include sequence-based features, as well as robustness to food processing, and address the issue of multiple protein isoforms. The latter is crucial in the development of a robust quantitative method but is seldom, if ever, considered in published studies. Eventually, 16 peanut peptide biomarkers were retained, covering all protein isoforms of Ara h 1 and Ara h 3, the two most abundant peanut allergens.

## 2. Material and methods

### 2.1. Reagent and materials

Urea, tris(hydroxymethyl)aminomethane (Tris-HCl), tris(2-carboxyethyl)phosphine hydrochloride (TCEP), iodoacetamide (IAA), tetraethylammonium bicarbonate (TEAB), and NAP-10 columns (GE Healthcare) were purchased from Sigma-Aldrich (Bornem, Belgium). Trifluoroacetic acid (TFA) was obtained from Thermo Fisher Scientific (Waltham, Massachusetts, USA), while Sep-Pak C18 1 cc Vac solid-phase extraction (SPE) cartridges were purchased from Waters (Milford, Massachusetts, USA). Enzymatic digestion was performed with MS-grade Trypsin Gold from Promega (Madison, Wisconsin, USA), and Lysyl Endopeptidase (Lys-C) was obtained from Wako Pure Chemical Industries, Ltd. (Osaka, Japan). Water, acetonitrile, methanol absolute, 2-propanol (ULC/MS-grade for all solvents), and formic acid were purchased from Biosolve (Valkenswaard, the Netherlands). ESI-L Low Concentration Tuning Mix was purchased from Agilent Technologies (Santa Clara, California, USA).

### 2.2. Food product preparation

Peanuts from two geographical regions and corresponding to two different market types were purchased from two providers. The first type of peanuts, provided by Quality Nuts BVBA (Zandhoven, Belgium), was cultivated in China and corresponds to the Spanish market type. The second type, provided by Versenoten.nl (Waddinxveen, The Netherlands), was cultivated in Israel and corresponds to the Virginia market type. Peanuts were purchased in an unprocessed, in-shell form and then shelled at the Food Pilot unit of the Flanders Research Institute for Agriculture, Fisheries, and Food (ILVO), producing five peanut matrices. Each peanut matrix was produced in duplicate, one for each peanut market type.

#### 2.2.1. Preparation of a raw peanut matrix

Raw, whole peanuts were ground using a Kenwood AT286 Glass mini chopper/mill attachment KW714229 (Kenwood, New Hampshire, UK) for 20 s on pulsing mode using the maximum speed, vacuum-packed (vacuum packaging machine KN5, vc999, Missouri, USA), and stored at 4 °C in the dark.

#### 2.2.2. Preparation of a heated peanut matrix

Raw, whole peanuts were roasted in batches of 500 g in an oven (baking oven condo CO 6.0608, Miwe, Germany) at 180 °C for 18 min. The peanuts were spread in a single layer on a baking plate and stirred with a spoon every 5 min to ensure homogenous roasting. The roasted peanuts were allowed to cool down at room temperature, ground into a peanut powder using ThermoResist Glazen Blender AT358 (Kenwood, New Hampshire, UK), vacuum-packed, and stored at 4 °C.

#### 2.2.3. Preparation of a low-pH peanut matrix: fermented peanut milk

This matrix was produced by fermenting peanut milk. Peanut milk was produced by soaking the peanuts in reverse osmose water overnight, followed by blending (Blixer 4VV -, 4.5 Liter, 1100 W, Variable Velocity: 300–3500 TPM, Robot Coupe, Vincennes Cedex, France) for 8 min at the maximum speed, resulting in a milky substance. The peanut milk was subsequently sieved (Retsch Test Sieve, 250 μm pore size, Retsch, Aartselaar, Belgium) and then transferred to a sterile Erlenmeyer flask of 1 L, and a bacterial culture (nu-trish® BY-Mild, which contains *Bifidobacterium* species, *Lactobacillus delbrueckii* subsp. *Bulgaricus,* and *Streptococcus thermophilus*, CHR Hansen, Hørsholm, Denemarken) was added (4 g of frozen culture for 600 mL of peanut milk). The flasks were aseptically sealed and incubated in a warm water bath (42 °C). Every 30 min, a sub-sample of the yogurt was checked for pH until a pH of 4.2 was reached. The yogurt was next divided into 20 g portions and stored at 4 °C in the dark. Given the lack of data concerning the stability of the produced yogurt, the analyses were performed within a month following the matrix production.

*2.2.4. Preparation of a matrix subjected to the maillard reactions: Caramelized peanuts*

An amount corresponding to 250 g of raw, whole peanuts was chopped (Blixer 4VV -, 4.5 Liter, 1100 W, Variable Velocity: 300–3500 TPM, Robot Coupe, Vincennes Cedex, France) 3 times for 5 s duration at the maximum speed, and 200 g lactose (D-Lactose monohydrate, Sigma-Aldrich, Bornem, Belgium) was mixed with 100 mL reverse osmose water to a smooth paste. The chopped peanuts were subsequently stirred into the lactose paste until all the peanut material was covered with the lactose paste. This peanut-lactose mixture was then spread over an oven plate and placed in a pre-heated (160 °C) convection oven for 15 min. Now, every 5 min, the mixture was stirred on the plate with a spoon to obtain homogeneous caramelization. The plate was left at room temperature to cool down for 16 h and, subsequently, ground into a fine powder (Kenwood Mini Chopper Mill Attachment AT320B). The caramelized peanut powder was divided into portions of 50 g, vacuum-packed (vacuum packaging machine KN5, vc999, Missouri, USA), and stored at 4 °C in the dark until use.

*2.2.5. Preparation of a complex and fatty-rich peanut matrix: Chocolate incurred with peanuts*

A peanut powder with particle sizes < 500 μm in diameter was produced by grinding raw peanuts (Kenwood Mini Chopper Mill Attachment AT320B, pulsing speed, 3 min, Kenwood, New Hampshire, UK), resulting in a slushy substance. This peanut mash was then defatted by pressing the material between extra-thick cotton cloths, resulting in dry peanut powder. The peanut mash and powder were weighed before and after defatting, and the loss in mass was documented to determine the amount of weight lost during defatting. The fine, dry peanut powder was then sieved with a Retsch test sieve (pore size 500 μm). Then, chocolate was prepared. First, a 'conche' was used to mix and smooth chocolate refiner flakes (90%) and cacao butter (10%). The refiner flakes (Callebaut, Lebbeke, Belgium) and cacao butter (Callebaut) were then warmed in a water bath in separate bowls until a temperature of 40–45 °C was reached. Next, the melted butter was slowly added to the warmed refiner flakes, and the mix was stirred for 3 min, followed by adding 1.2% ammonium phosphatide. This mixture was stirred for an additional 3 min. Finally, the chocolate incurred with peanuts (38,178 ppm [mg peanut proteins/kg chocolate] for Spanish and 36,792 ppm for Virginia market type peanuts) was produced by adding 70 g of defatted peanut powder to the liquid chocolate and stirring for 3 min while maintaining the temperature of the chocolate at 42 °C. The liquid chocolate was subsequently poured into sterile chocolate molds, covered with plastic foil, and left to cool and solidify overnight at 4 °C. The following day, the chocolate bars were rasped into chocolate flakes, which were then separated into aliquots of 50 g and stored at 4 °C in the dark. A proper control/blank chocolate material was prepared without peanut powder.

*2.3. Sample preparation, mass spectrometry analysis and data treatment*

The protocol used to prepare the samples is comparable to the method described in our previous study dedicated to eggs (Gavage et al., 2018). Briefly, proteins were extracted from food matrices, purified, and proteolytically digested. Two digestion protocols were compared for each processed food product. In the first protocol, the proteins were digested with trypsin alone. In the second protocol, tandem endoproteinase Lys-C/trypsin digestion was used. For each combination of digestion protocol and processed food product, three biological replicates were prepared and analyzed via HPLC-HRMS.

Mass spectrometry analysis and obtained data treatment were also performed as described in our previous article. For each sample, 3 μg of peptides were analyzed using an ESI-MS/MS maXis Impact UHR-TOF (Bruker, Billerica, Massachusetts, USA) coupled to a Dionex UltiMate 3000 Standard LC Systems (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Peptides were separated by reverse-phase liquid chromatography on an Acclaim PepMap100 C18 column (3 μm, 100 Å, 1 mm × 15 cm nanoViper, Thermo Fisher Scientific, Waltham, Massachusetts, USA) with a 60 min gradient elution with a mixture of water/acetonitrile, each containing 0.1% formic acid. The data were processed using Mascot 2.5 (Matrix Science, London, United Kingdom) as the search engine and a homemade protein database, adapted to peanut samples, and downloaded from UniProt (uniprot.org). This database contains the Arachis genus proteome (which includes peanuts [Arachis hypogaea]) and some possible contaminants (i.e., cacao, bacteria from Nu-trish® BY-Mild [Bifidobacterium animalis subsp. lactis, Lactobacillus delbrueckii subsp. Bulgaricus, and Streptococcus thermophilus] used for yogurt preparation, human keratin, trypsin, and Lys-C) to avoid misidentification. In total, this database contains 78,112 entries (1846 entries for the Arachis genus [1492 for peanut Arachis hypogaea], 34,951 for bacteria from Nu-trish® BY-Mild, 40,956 for cacao, and 350 for human keratin). Peptide and protein identifications were validated using Scaffold 4.8 (Proteome Software, Portland, Oregon, USA). Combination of the list of validated peptides and buckets generated by ProfileAnalysis (Bruker, Billerica, Massachusetts, USA) software allowed the generation of comparable results. Findings discussed below are based on signal intensities linked to the identified peptides and were obtained using this method.

*2.4. Multiple sequence alignments and principal component analysis*

Multiple sequence alignments were performed on the UniProt website (uniprot.org/align/) using Clustal Omega software (Sievers et al., 2011). Obtained alignments were transferred to Jalview software (Waterhouse, Procter, Martin, Clamp, & Barton, 2009) for analysis. Jalview was notably used to perform principal component analysis (PCA) on protein sequence alignments by selecting the BLOSUM62 scoring matrix.

## 3. Results

*3.1. General strategy*

Food allergies are an increasing problem not only for allergic patients and their social networks but also for the food industry and regulators. Undeclared allergens are, indeed, the primary cause of food product recalls and alerts in Western countries (Bucchini, Guzzon, Poms, & Senyuva, 2016). Therefore, an urgent need exists for reference methods to detect and quantify allergens in food products (Walker, Burns, Elliott, Gowland, & Mills, 2016).

The objective of this study was to identify potential peanut peptide biomarkers using an HRMS-based discovery proteomics approach. More than 300 peanut peptides were identified in the analyzed samples (Table 1), and a complete list of these peptides and their intensity in each peanut matrix can be found in Supplementary Material 1. The potential peanut peptide biomarkers were then used to develop a UHPLC-MS/MS method for quantifying peanuts in processed food products in routine settings.

Samples were prepared according to the procedure developed in a previous study for the identification of peptide biomarkers in eggs (Gavage et al., 2018). In that study, some egg proteins, such as ovalbumin, were found to be resistant to proteolytic digestion using a trypsin-only protocol. Therefore, we implemented a digestion protocol combining two enzymes, endoproteinase Lys-C and trypsin, which improved the digestion of resistant proteins as evident from the significantly higher MS signal intensities of peptides originating from such proteins. The two digestion protocols (i.e., trypsin-only and tandem Lys-C/trypsin) were applied in parallel to all peanut samples. Unlike egg proteins, no proteins appeared to be resistant to proteolytic digestion using only trypsin as only minor differences were observed on the peptide signal intensities when peanut samples digested with trypsin alone were compared to those digested with tandem Lys-C/trypsin. The

**Table 1**
Unique proteins and number of unique peptides for these proteins identified following HRMS analysis of peanut samples (raw peanuts, heated peanuts, low pH matrix, caramelized peanuts and fatty rich environment).

| Identified protein | UniProt entry identifier(s) | Number of identified unique peptides | Coverage |
|---|---|---|---|
| Ara h 1 | B3IXL2, E5G076, N1NG13, P43237, P43238, Q6PSU3 | 67 | 61.4% |
| Ara h 2 | Q6PSU2 | 12 | 41.9% |
| Ara h 3 | A1DZF0, B5TYU1, E5G077, Q0GM57, Q5I6T2, Q6T2T4, Q647H2, Q647H3, Q647H4, Q8LKN1, Q9FZ11, Q9SQH7 | 127 | 83.2% |
| Ara h 6 | A1DZE9 | 5 | 24.8% |
| Ara h 8 | B0YIU5, B1PYZ4, Q6VT83 | 6 | 31.4% |
| Ara h 10 | Q647G5 | 3 | 23.1% |
| Ara h 13 | C0HJZ1 | 4 | 36.1% |
| 11S arachin | E9LFE8 | 7 | 30.8% |
| 11S seed storage globulin B1 | A1E2B0 | 6 | 25.9% |
| Annexin | A0A0F6VX63, F6KLJ6 | 7 | 23.2% |
| Bowman-Birk trypsin inhibitor | P01066, Q0PKR5, Q7X973 | 2 | 40% |
| Conarachin | Q647H1 | 10 | 33.9% |
| Galactose-binding lectin | A0A089ZXL7, Q38711 | 14 | 55% |
| Glyceraldehyde-3-phosphate dehydrogenase C2 | A0A0A6ZDP1, A0A0A6ZDT0 | 4 | 38.2% |
| Late embryogenesis abundant protein 3 | A0A1L5JJI4 | 2 | 7.0% |
| Late embryogenesis abundant protein group 5 protein | E5FHZ2 | 5 | 21.1% |
| Lipoxygenase | Q4JME6, Q4JME7 | 24 | 27.2% |
| Ribulose bisphosphate carboxylase large chain | A0A191UJ50, O20356 | 4 | 11.8% |
| Seed maturation protein | N1NKG9 | 2 | 23.4% |
| Steroleosin-A | A7LB60 | 6 | 16.6% |
| Steroleosin-B | A7LB59 | 8 | 18.1% |

results discussed below in further detail correspond to samples digested with trypsin alone; however, both digestion protocols' results are included in Supplementary Material 1.

To ensure the robustness, specificity, sensitivity and trueness of a future quantitative method, selected peptide biomarkers have to fulfill several criteria. The peptides have to be specific to peanuts, robust to food processing, derived from abundant proteins, size-compatible with *m/z* analyzers (i.e., 8–25 amino acids), not prone to missed cleavages by the protease(s) used or to amino acid modifications, and present at comparable levels regardless of the peanut market type. Thus, to identify ideal peanut peptide biomarkers, the detected peptides were further filtered using these criteria as detailed in the following sections.

### 3.2. Peanut protein composition

Peanuts are seeds and, therefore, the vast majority of proteins recovered/extracted from them are seed storage proteins. These proteins serve as a reserve of amino acids for the growing embryo upon germination. The most relevant peanut allergens are storage proteins, including the four most abundant: Ara h 1 (17.1 ± 3.4% of total extracted protein), Ara h 2 (6.2 ± 1.3% of total extracted protein), Ara h 3 (70.6 ± 8.6% of total extracted protein) and Ara h 6 (5.8 ± 1.8% of total extracted protein) with an extraction yield of 74.7 ± 8.5% for the 20 analyzed peanut samples (Koppelman et al., 2016). As the ultimate aim of this study was to develop a sensitive, quantitative UHPLC-MS/MS-based detection method, the method, logically, must target these abundant proteins. This aspect was demonstrated here because, in the 60 analyses performed (i.e., 5 peanut food products, 2 market types, 2 digestion protocols, and 3 biological replicates), over 95% of the total MS signal intensities measured are associated with identified peanut peptides belonging to these 4 proteins. In fact, Ara h 1 peptides accounted for 21% of the signal, while Ara h 2, Ara h 3, and Ara h 6 peptides accounted for 1%, 74%, and 0.5%, respectively. Given these results, only peptides from Ara h 1 and Ara h 3 were further considered as potential peptide biomarkers for the remainder of this study.

### 3.3. Peptide sequence-based criteria

With the goal of developing a quantitative UHPLC-MS/MS method, total peanut protein will be quantified through peanut peptides. To ensure method trueness, the quantification of these peptides should directly correlate with the protein amount. The ideal case would be a stoichiometric ratio between a given peptide and its corresponding protein, for example, *x* mole(s) of a given peptide corresponding to × mole(s) of its corresponding protein. To achieve this goal, peptides carrying amino acids that can potentially be modified or are subject to incomplete enzymatic digestion should be discarded (Gavage et al., 2018).

Potential amino acid modifications were first evaluated using a literature review (Reimer et al., 2011) and then empirically confirmed using Mascot's 'error-tolerant search'. The observed modifications are cysteine carbamidomethylation, methionine oxidation, *N*-terminal glutamine to pyroglutamate conversion and, asparagine and glutamine deamidation. Cysteine carbamidomethylation, inevitably, results from the reduction and alkylation step during sample preparation using TCEP and IAA. The other modifications are spontaneous and uncontrolled reactions. Therefore, peptides containing cysteine, methionine, or *N*-terminal glutamine were systematically discarded. Deamidation was treated differently, given the high occurrence of asparagines and glutamines (Andersen, Hill, Gorbet, & Brodbeck, 1998). Although both amino acids are prone to deamidation, the rate of glutamine deamidation was reported to be much lower than that of asparagine (Robinson et al., 2004). Glutamine deamidation was, therefore, not further considered. Moreover, no experimental evidence of glutamine deamidation was observed in our data. Deamidation is influenced by the nature of the amino acid at the carboxyl side of asparagine (Asn). The fastest reaction occurs when this amino acid is glycine (Gly), whereas for other amino acids, the reaction rates are at least 10-times lower (Kosky, Dharmavaram, Ratnaswamy, & Manning, 2009; Robinson & Robinson, 2004). Thus, only peptides containing Asn-Gly were discarded. Finally, post-translational modifications documented in databases, such as UniProt, were also considered. Glycosylation on $Asn^{521}$ is, for example, documented for Ara h 1. We decided to reject peptides holding this glycosylation site to prevent that incomplete post-translational modification would negatively influence further results.

The types of amino acids surrounding the proteolytic recognition site are known to affect the efficiency of peptide bond hydrolysis (Siepen, Keevil, Knight, & Hubbard, 2007). Sequences that have a

negative effect on this efficiency must, therefore, be avoided. Protease recognition sites are described using the Schechter and Berger nomenclature (Schechter & Berger, 1967), in which amino acids surrounding the cleavage site are indicated as P4-P3-P2-P1-P1′-P2′-P3′, where the peptide bond between P1 and P1′ is the cleaved bond. Arginine, lysine, and proline negatively affect digestion when present at P1′. Peptides with one of these amino acids in P1′ were also discarded. The acidic amino acids aspartate and glutamate also negatively influence digestion when they are present near the cleavage site. Given the relatively high occurrence of such amino acids, systematically discarding possibly-affected peptides is too restrictive. Peptides containing aspartate or glutamate close to the cleavage site were treated on a case-by-case basis, evaluating the importance of the potentially-missed cleavage. A peptide was rejected when the MS signal intensity of the associated peptide containing the missed cleavage exceeded 10% of the MS signal of the fully tryptic peptide.

### 3.4. Robustness to food processing

Peanuts are consumed in many forms, such as boiled, roasted, or added to snack foods. However, food processing and matrix interactions are known to affect the recovery and subsequent detectability of targeted compounds (Khuda, Jackson, Fu, & Williams, 2015; Sayers et al., 2016). To ensure the accuracy and robustness of our quantitative UHPLC-MS/MS method, the effects of food processing were considered when selecting peptide biomarkers. Matrices and processing conditions applied to peanut samples were selected to represent the conditions of daily food products. Selected peptide biomarkers should, ideally, not be affected in these conditions.

Matrix and food processing effects were evaluated comparing the MS signal intensities of the identified peptides under the five processing conditions assessed. Unaffected peptides should have comparable MS intensities in different conditions. Results for Ara h 1 and Ara h 3 are shown in Fig. 1. Only peptides fulfilling sequence-based criteria are displayed in this figure for both peanut market types, Spanish and Virginia. Some peptides, such as GTGNLELVAVR from Ara h 1 and SQSENFEYVAFK from Ara h 3, seem to be largely resistant to the food processes selected in this study (i.e., heat treatment, caramelization, a low-pH matrix, and a complex matrix). Conversely, the intensity of some peptides depends on the food processing method, with significantly lower MS intensities under some conditions. For example, the intensity of the peptide IPSGFISYILNR from Ara h 1 was significantly lower in samples corresponding to the heat treatment, low pH matrix, and complex matrix (chocolate) conditions when compared to unprocessed samples (Student's *t*-test, $p < 0.05$). Potential peptide biomarkers have, thus, been refined based on these results as the detection of the selected peptides must be robust, even when various food processing methods are used, to ensure the future quantitative UHPLC-MS/MS method is robust.

### 3.5. Market type variation and protein isoforms

Differences in protein content and protein relative-abundance were shown to vary among the tested peanut varieties (Kottapalli et al., 2008; Schmidt et al., 2009). Moreover, for some allergens, Ara h corresponds to multiple protein isoforms with sequence variations, and the relative abundance of the different allergen isoforms is also likely to vary depending on the variety (Jin, Guo, Chen, Howard, & Zhang, 2009). To consider this potential variability in our peptide biomarkers selection, peanuts from two market types (i.e., Spanish and Virginia) were analyzed.

Peanut detection methods identified in other studies have rarely included aspects of market type variation and protein isoforms (Johnson et al., 2016). However, this additional degree of complexity must be taken into account when developing absolute quantification methods. Indeed, proposed thresholds for food allergens are expressed as total protein quantities. Based on these thresholds, or reference doses (e.g., those proposed by VITAL® 2.0), most (i.e., 95–99%) of the allergic population should be protected, and appropriate measures may be taken when undeclared allergens are detected. Peptide-based approaches, therefore, require a conversion factor to translate results expressed in peptide quantities into total protein amounts. This conversion factor undeniably needs to include the proportion of the protein corresponding to the targeted peptide, compared to the total ingredient protein. Variability, in the proportion of a protein of interest based on the market type, increased the quantification method's inaccuracy in this study. Ideal peptide biomarkers should, thus, be present at a constant concentration, irrespective of the peanut variety, and include all protein isoforms.

Possible variation due to the peanut's origin was also evaluated by comparing peptide MS intensities. A comparison of the peptides fulfilling the sequence-based criteria is presented in Fig. 2. No significant differences are observed for peptides from Ara h 1, but the results are more complex for Ara h 3. While some peptides, TANDLNLLILR and WLGLSAEYGNLYR, have comparable intensities when the two market types are compared, other peptides are present in higher levels in one of the market types. This was the case for SQSENFEYVAFK and VYDEELQEGHVLVVPQNFAVAGK, for which peak intensities are higher in the Spanish market type, whereas the peak intensities of other peptides (e.g., FYLAGNQEQEFLR and VYDEELQEGHVLVVPQNFAVAAK) are more intense in the Virginia market type.

Allergens Ara h 1 and Ara h 3 do not necessarily correspond to unique proteins but, rather, to groups of protein isoforms. The UniProt database contains six entries (i.e., B3IXL2, E5G076, N1NG13, P43237, P43238, and Q6PSU3) corresponding to Ara h 1. These entries correspond to three isoforms as some have identical protein sequences (e.g., B3IXL2/P43237/Q6PSU3 and N1NG13/P43238). Concerning Ara h 3, 12 entries corresponding to 12 isoforms with different protein sequences can be found. The protein sequences of the different isoforms of Ara h 1 and Ara h 3 are not fully identical. Some sequence-based selected peptides were found to be specific to one or several of these isoforms. These peptides' specificity for the different isoforms is shown in Table 2 for Ara h 1 and Ara h 3. Some of the sequence-based selected peptides of Ara h 1, DQSSYLQGFSR and GTGNLELVAVR, are common to all three Ara h 1 isoforms. However, other peptides, such as EHVQELTK and SSENNEGVIVK, are specific to one or two of these isoforms. As previously mentioned, no significant differences in peptide MS signal intensity were observed between the two peanut market types, which means the abundance of Ara h 1 is comparable in Spanish and Virginia market type peanuts, considering the samples used in this study. Moreover, the abundance of each isoform seems to be similar in both market types.

Given the higher number of isoforms, the case of Ara h 3 is more complex. None of the 32 sequence-based selected peptides are common to all Ara h3 isoforms. Thus, to cover all these isoforms and quantify the total abundance of this allergen, multiple peptides must be considered. The different isoforms were gathered in groups based on sequence similarities. Principal component analysis (PCA) was performed on the sequence alignment of the 12 protein isoforms of Ara h 3, which generated a spatial representation of the sequences' similarities. The two-dimensional PCA (Fig. 3) indicates the presence of three isoform groups arbitrarily called Group 1, Group 2 and Group 3. The selection of peptides should, therefore, contain peptides from each group. These 3 groups can be correlated to the variation observed in Fig. 2. Indeed, since the sequences of different isoforms of Ara h 3 are not fully identical, some peptides are specific to particular isoforms or groups of isoforms. The complete peptide distribution among the different isoforms and groups is displayed in Table 2.

Of the 32 selected peptides of Ara h 3, 5 peptides are specific to Group 3, which is composed of the sole Q647H2 isoform of Ara h 3. These peptides display weak MS signal intensities when compared to the intensity of other Ara h 3 peptides, such as AQSENYEYLAFK and

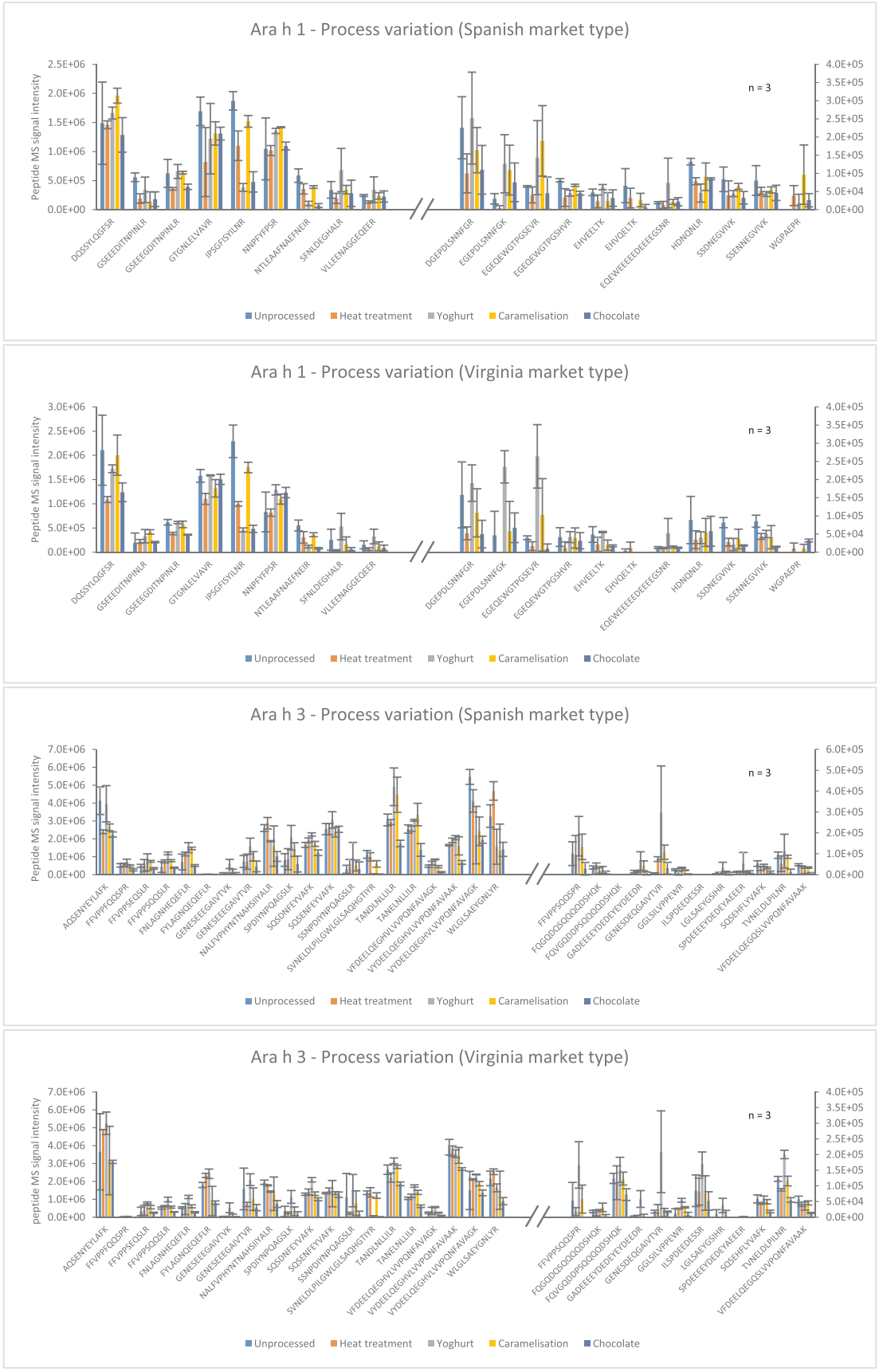



**Fig. 1.** Effect of food processing on Ara h 1 and Ara h 3. Exposed peptides fulfil the sequence based criteria. For each market type (Spanish and Virginia), five peanut matrices are compared. From left to right: no process, heated peanut, low pH matrix (yoghurt), caramelized peanut and fatty rich and complex environment (chocolate). The comparison is based on the MS signal intensity. The results are expressed as the means of the peptide MS signal intensity ± 1 SD for 3 independent replicates. In each histogram, peptides are separated in two intensity-based groups with different axis scales to improve low intensity peptides visualisation.

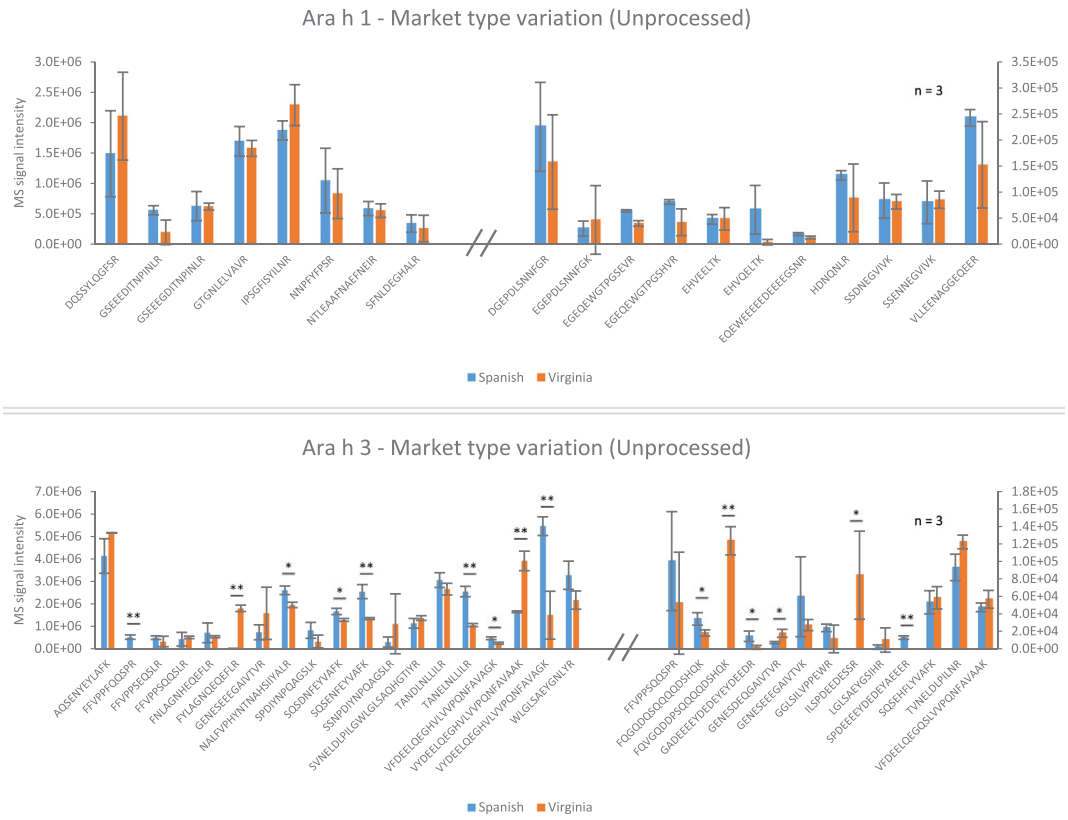



**Fig. 2.** Peptide MS signal intensity comparison between the two peanut market types (Spanish and Virginia) for Ara h 1 and Ara h 3. Only peptides fulfilling sequence based criteria were selected. Exposed results correspond to analysis of unprocessed peanuts and are expressed as the means of the peptide MS signal intensity ± 1 SD for 3 independent replicates. In each histogram, peptides are separated in two intensity-based groups with different axis scales to improve low intensity peptides visualisation. Statistical significance is evaluated with Student's T-test (* = $p < 0.05$ and ** = $p < 0.01$).

TANDLNLLILR, which may be due to a lower abundance of this isoform or, potentially, to technical effects (e.g., incomplete digestion or poor ionization). Given their very low MS intensities, these peptides were not considered as interesting biomarkers. However, the Q647H2 isoform of Ara h 3 could be present at higher levels in other peanut varieties or under different growth conditions. Therefore, at least one peptide of this isoform was considered in our selection. Based on its robustness to food processing and its MS signal intensity, TVNELDLPILNR seems the best candidate to represent this isoform by our selected peptides.

MS signal intensities associated with peptides from Group 2 seem to be higher in the Virginia market type samples. This Group 2 is composed of 2 protein isoforms, and 7 sequence-based selected peptides are specific to them. Among these peptides, 4 have a significantly higher MS signal intensity or are only detected in samples from the Virginia market type. In this case, since samples were prepared with the same protocol, the only different parameter is the peanut market type. It could, therefore, be concluded that protein isoforms from Group 2 are more abundant in Virginia market type peanuts.

Group 1 corresponds to 9 protein isoforms and 20 sequence-based selected peptides specific to this group. Two peptides (i.e., FNLAGNHEQEFLR and WLGLSAEYGNLYR) are found in all 9 isoforms, whereas others are specific to protein isoforms, display equivalent MS signal intensities in both market types (e.g., FNLAGNHEQEFLR and TANDLNLLILR), or display a higher intensity in Spanish market type peanut samples (e.g., TANELNLLILR and VYDEELQEGHVLVVPQNFAVAGK). None of the peptides in Group 1 are significantly more intense in Virginia market type samples. It can, therefore, be concluded that some protein isoforms of Group 1 are more abundant in Spanish market type peanuts.

Based on these considerations, peptides from Ara h 1 seem to be the best potential biomarkers. Indeed, no variation between the two considered market types was observed. Moreover, the number of isoforms is limited, and some sequence-based selected peptides are well conserved in all isoforms. However, Ara h 3 is the most abundant protein in peanuts and, hence, a suitable target to ensure high sensitivity in a quantitative method. A combination of peptides could allow coverage of all Ara h 3 isoforms, as well as the global quantification of Ara h 3, so at this stage, peptide biomarker selection should not be limited to Ara h 1 peptides. Because of the high abundance of Ara h 3, peptides from this allergen were also selected despite the observed variabilities.

At this stage, considering sequence-based criteria, robustness to food processing, and the multiple isoforms, 17 peptides were pre-selected (i.e., 5 from Ara h 1 and 12 from Ara h 3). With this pre-selection, all isoforms of both Ara h 1 and Ara h 3 are covered, ensuring method robustness. The final selection of biomarkers will be performed during the development of the quantitative UHPLC-MS/MS method.

### *3.6. Specificity of selected peptide biomarkers and comparison with the literature*

Via an empirical approach using HRMS, more than 300 peanut peptides were identified in the various peanut-based test materials. Applying the criteria extensively described in the previous sections, these peptides were refined to obtain a selection of potentially salient peptide biomarkers in peanuts. The last step involves assessing the selected peptides' specificity to peanuts. This verification was performed using basic local alignment search tool (BLAST) analyses considering the UniProt and NCBI protein databases. The selected peptides must be specific to peanuts and not found in the protein sequences of other food ingredients because, otherwise, false-positive results could occur.

**Table 2**
Specificity of the sequence based selected peptides to the different protein isoforms of Ara h 1 and Ara h 3 identified in UniProt database (uniprot.org). Isoforms of Ara h 3 are separated in the 3 artificial groups highlighted by the PCA performed on the protein sequences alignment.

| Ara h 1 peptides | B3IXL2/P43237/Q6PSU3 | N1NG13/P43238 | E5G076 |
|---|---|---|---|
| DGEPDLSNNFGR | X | | |
| DQSSYLQGFSR | X | X | X |
| EGEPDLSNNFGK | | X | |
| EGEQEWGTPGSEVR | X | | |
| EGEQEWGTPGSHVR | | X | X |
| EHVEELTK | | X | X |
| EHVQELTK | X | | |
| EQEWEEEEEDEEEEGSNR | X | | X |
| GSEEEDITNPINLR | X | | |
| GSEEEGDITNPINLR | | X | X |
| GTGNLELVAVR | X | X | X |
| HDNQNLR | X | X | |
| IPSGFISYILNR | X | X | |
| NNPFYFPSR | X | X | X |
| NTLEAAFNAEFNEIR | X | X | X |
| SFNLDEGHALR | X | X | |
| SSDNEGVIVK | X | | |
| SSENNEGVIVK | | X | |
| VLLEENAGGEQEER | X | X | X |
| WGPAEPR | X | | |

| Ara h 3 peptides | Group 1 | | | | | | | | | Group 2 | | Group 3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A1DZF0 | B5TYU1 | Q5I6T2 | Q6T2T4 | Q647H3 | Q647H4 | Q8LKN1 | Q9FZ11 | Q9SQH7 | E5G077 | Q0GM57 | Q647H2 |
| AQSENYEYLAFK | | | | | | | | | | X | X | |
| FFVPPFQQSPR | X | | | | | | | | X | | | |
| FFVPPSEQSLR | | | | X | | X | X | | | | | |
| FFVPPSQQSLR | | | | | | | | X | | | | |
| FFVPPSQQSPR | | X | X | | | | | | | | | |
| FNLAGNHEQEFLR | X | X | X | X | X | X | X | X | X | | | |
| FQGQDQSQQQQDSHQK | | | | X | | X | X | | | | | |
| FQVGQDDPSQQQQDSHQK | | | | | | | | | | X | X | |
| FYLAGNQEQEFLR | | | | | | | | | | X | X | |
| GADEEEEYDEDEYEYDEEDR | | X | X | | | | | X | | | | |
| GENESDEQGAIVTVR | | | | X | X | X | X | | | | | |
| GENESEEEGAIVTVK | | | | | | | | X | | | | |
| GENESEEEGAIVTVR | | | X | | | | | | X | | | |
| GGLSILVPPEWR | | | | | | | | | | | | X |
| ILSPDEEDESSR | | | | | | | | | | X | X | |
| LGLSAEYGSIHR | | | | | | | | | | | | X |
| NALFVPHYNTNAHSIIYALR | X | X | X | X | X | X | X | X | X | | | |
| SPDEEEEYDEDEYAEEER | X | | | | | | | | | | | |
| SPDIYNPQAGSLK | X | X | X | X | X | X | X | X | | | | |
| SQSDNFEYVAFK | | X | X | | X | | | X | | | | |
| SQSEHFLYVAFK | | | | | | | | | | | | X |
| SQSENFEYVAFK | X | | | X | | X | X | | X | | | |
| SSNPDIYNPQAGSLR | | | | | | | | | | X | X | |
| SVNELDLPILGWLGLSAQHGTIYR | | | | | | | | | | X | X | |
| TANDLNLLILR | | | X | | X | | | X | | | | |
| TANELNLLILR | X | X | | X | | X | | | | | | |
| TVNELDLPILNR | | | | | | | | | | | | X |
| VFDEELQEGHVLVVPQNFAVAGK | | | | X | | X | X | | | | | |
| VFDEELQEGQSLVVPQNFAVAAK | | | | | | | | | | | | X |
| VYDEELQEGHVLVVPQNFAVAAK | | | | | | | | | | X | X | |
| VYDEELQEGHVLVVPQNFAVAGK | X | X | X | | X | | | X | X | | | |
| WLGLSAEYGNLYR | X | X | X | X | X | X | X | X | X | | | |

Among the 17 pre-selected peptides, only the peptide FYLAGNQEQEFLR from Ara h 3 was rejected based on BLAST results as this peptide is also present in faba beans (*Vicia faba* var. *minor*) protein. Since this bean is used not only in animal feed but also in human food (Crépon et al., 2010; Smith & Hardacre, 2011), this peptide was rejected because of its lack of specificity to peanuts. The final selected peptide biomarkers (Table 3) contain 16 peptides from the two most abundant proteins in peanuts, Ara h 1 (5 peptides) and Ara h 3 (11 peptides). In the future, these peptides will be used to develop a UHPLC-MS/MS routine method to quantify peanuts in processed food products.

The 16 selected peanut peptides were, finally, compared with biomarkers described in the literature. The literature review is summarized in Supplementary Material 2, where our potential peptide biomarkers are highlighted in bold, and the matrices used in the different papers are indicated. A total of 30 peanut biomarkers were identified from the literature, and most of them (28 out of 30) are from Ara h 1, Ara h 2, or Ara h 3. Some of our selected peptides, such as GTGNLELVAVR (Ara h 1) and WLGLSAEYGNLYR (Ara h 3), appear to be recurrent biomarkers. However, some cited peptides do not meet our selection criteria and were not considered as potential biomarkers in this study. For example, peptide CMCEALQQIMENQSDR, which was found in 3 publications, contains potentially modified cysteines and methionines. Furthermore, some of our selected peptides (3 from Ara h 1 and 3 from Ara h 3) were not identified as biomarkers in the considered literature. These peptides

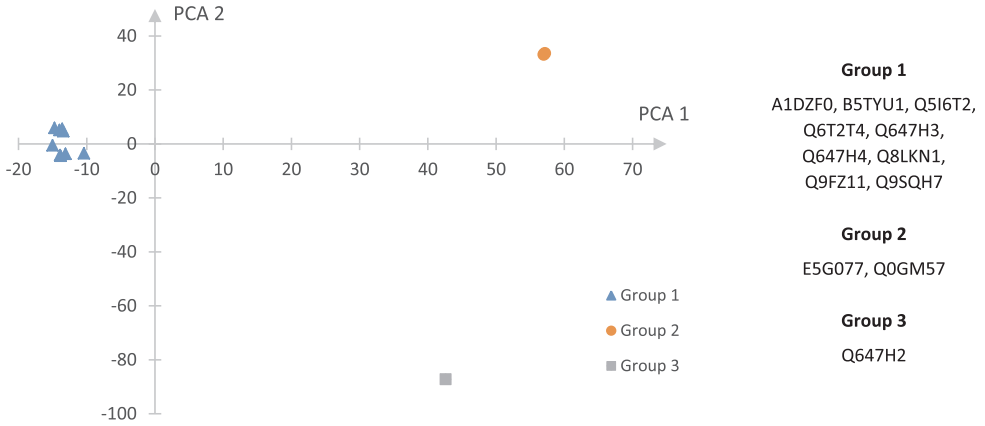



**Fig. 3.** Two dimensional principal component analysis (PCA) performed on the protein sequence alignment of the 12 isoforms of Ara h 3 found in the UniProt database. PCA highlight the presence of 3 groups of isoforms with close protein sequences and called Group 1, Group 2 and Group 3.

**Table 3**
Final potential peptide biomarkers selection fulfilling the above-mentioned criteria. The peptides are linked to their respective allergen (Ara h 1 and Ara h 3).

| Allergen | Selected peanut peptide biomarkers |
|---|---|
| Ara h 1 | DQSSYLQGFSR |
| | GSEEEDITNPINLR |
| | GSEEEGDITNPINLR |
| | GTGNLELVAVR |
| | NNPFYFPSR |
| Ara h 3 | AQSENYEYLAFK |
| | FNLAGNHEQEFLR |
| | NALFVPHYNTNAHSIIYALR |
| | SQSDNFEYVAFK |
| | SQSENFEYVAFK |
| | TANDLNLLILR |
| | TANELNLLILR |
| | TVNELDLPILNR |
| | VYDEELQEGHVLVVPQNFAVAAK |
| | VYDEELQEGHVLVVPQNFAVAGK |
| | WLGLSAEYGNLYR |

are DQSSYLQGFSR, GSEEEDITNPINLR, and GSEEEGDITNPINLR for Ara h 1, and NALFVPHYNTNAHSIIYALR, SQSDNFEYVAFK, and TVNELDLPILNR for Ara h 3. Hence, they could be considered as new peptide biomarkers for peanuts as they meet our established selection criteria. We noticed that none of the 15 Ara h 3 peptide biomarkers identified in the literature are present in the Q647H2 Ara h 3 isoform sequence. This omission could lead to bias and total Ara h 3 under-estimation in a future quantitative method.

## 4. Conclusions

An allergic reaction to peanuts is one of the most serious types of food allergies, considering their prevalence and the potential severity of this reaction, which can lead to anaphylactic shock. To provide more protection for allergic patients via improved food labeling, accurate detection and quantification methods are needed for peanuts and, more broadly, all food allergens. Mass spectrometry-based methods are promising and can overcome the limitations of existing methods.

Identifying potential allergen biomarkers is the first step in the development of a mass spectrometry-based method. Using an empirical approach based on high-resolution mass spectrometry, this study presents the identification of potential peptide biomarkers to detect and quantify peanuts in processed food.

More than 300 peptides were identified during the analysis of various processed peanut matrices (i.e., raw peanuts, heated peanuts, a low-pH peanut matrix, caramelized peanuts, and peanuts in a fat-rich environment). These peptides were filtered using a set of selection criteria to ensure method specificity, sensitivity, trueness, and robustness. Ideal peptide biomarkers must be specific to peanuts, belong to abundant proteins, and be robust to food processing but not prone to missed cleavages by the protease used or to amino acid modifications. Multiple protein isoforms and origin variation issues were also considered by analyzing peanuts from two geographical regions corresponding to two market types. Ara h 1 includes 3 isoforms with relatively close sequences and several common tryptic peptides. No significant differences were observed between the two market types (i.e., Virginia and Spanish) concerning the amount of Ara h 1. However, Ara h 3 corresponds to 12 protein isoforms with more significant sequence variations, and no tryptic peptide was found conserved between these 12 isoforms. The abundance of some Ara h 3 isoforms was found to vary by peanut market type. These aspects are hardly considered in peanut peptide biomarkers identification studies or in detection and quantification method development. They are, however, essential to ensure the robustness of the corresponding method. Our selection includes peptides covering all protein isoforms of Ara h 1 and Ara h 3. Our approach led to the identification of 16 potential peanut peptide biomarkers. These peptides will be used as targets in the future development of a routine UHPLC-MS/MS method to detect and quantify allergens in processed food products.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

The research that yielded these results was funded by the Belgian Federal Public Service of Health, Food Chain Safety and Environment through the contract RT 15/10 ALLERSENS. The authors also thank MaSUN, the Mass Spectrometry facility of the University of Namur (URBC-Narilis).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2019.125428.

## References

Andersen, P. C., Hill, K., Gorbet, D. W., & Brodbeck, B. V. (1998). Fatty acid and amino acid profiles of selected peanut cultivars and breeding lines. *Journal of Food*

*Composition and Analysis, 11*(2), 100–111. https://doi.org/10.1006/JFCA.1998.0565.
Bégin, P., Winterroth, L. C., Dominguez, T., Wilson, S. P., Bacal, L., Mehrotra, A., ... Nadeau, K. C. (2014). Safety and feasibility of oral immunotherapy to multiple allergens for food allergy. *Allergy, Asthma & Clinical Immunology, 10*(1), 1. https://doi.org/10.1186/1710-1492-10-1.
Boo, C. C., Parker, C. H., & Jackson, L. S. (2018). A targeted LC-MS/MS method for the simultaneous detection and quantitation of egg, milk, and peanut allergens in sugar cookies. *Journal of AOAC International, 101*(1), 108–117. https://doi.org/10.5740/jaoacint.17-0400.
Bucchini, L., Guzzon, A., Poms, R., & Senyuva, H. (2016). Analysis and critical comparison of food allergen recalls from the European Union, USA, Canada, Hong Kong, Australia and New Zealand. *Food Additives and Contaminants - Part A Chemistry, Analysis, Control, Exposure and Risk Assessment, 33*(5), 760–771. https://doi.org/10.1080/19440049.2016.1169444.
Crépon, K., Marget, P., Peyronnet, C., Carrouée, B., Arese, P., & Duc, G. (2010). Nutritional value of faba bean (Vicia faba L.) seeds for feed and food. *Field Crops Research, 115*, 329–339. https://doi.org/10.1016/j.fcr.2009.09.016.
Croote, D., & Quake, S. R. (2016). Food allergen detection by mass spectrometry: The role of systems biology. *NPJ Systems Biology and Applications, 2*, 16022. https://doi.org/10.1038/npjsba.2016.22.
De Silva, D., Geromi, M., Panesar, S. S., Muraro, A., Werfel, T., Hoffmann-Sommergruber, K., ... Sheikh, A. (2014). Acute and long-term management of food allergy: Systematic review. *Allergy: European Journal of Allergy and Clinical Immunology, 69*, 159–167. https://doi.org/10.1111/all.12314.
Gavage, M., Van Vlierberghe, K., Van Poucke, C., De Loose, M., Gevaert, K., Dieu, M., ... Gillard, N. (2018). Selection of egg peptide biomarkers in processed food products by high resolution mass spectrometry. *Journal of Chromatography A*. https://doi.org/10.1016/J.CHROMA.2018.11.036.
Henrottin, J., Planque, M., Huet, A.-C., Marega, R., Lamote, A., & Gillard, N. (2019). Gluten analysis in processed foodstuffs by a multi-allergens and grain-specific UHPLC-MS/MS method: One method to detect them all. *Journal of AOAC International*. https://doi.org/10.5740/jaoacint.19-0057.
Husain, Z., & Schwartz, R. A. (2012). Peanut allergy: An increasingly common life-threatening disorder. *Journal of the American Academy of Dermatology, 66*(1), 136–143. https://doi.org/10.1016/J.JAAD.2011.02.031.
Jin, T., Guo, F., Chen, Y. wei., Howard, A., & Zhang, Y. Z. (2009). Crystal structure of Ara h 3, a major allergen in peanut. *Molecular Immunology, 46*(8–9), 1796–1804. https://doi.org/10.1016/j.molimm.2009.01.023.
Johnson, P. E., Sayers, R. L., Gethings, L. A., Balasundaram, A., Marsh, J. T., Langridge, J. I., & Mills, E. N. C. (2016). Quantitative proteomic profiling of peanut allergens in food ingredients used for oral food challenges. *Analytical Chemistry, 88*(11), 5689–5695. https://doi.org/10.1021/acs.analchem.5b04466.
Khuda, S. E., Jackson, L. S., Fu, T. J., & Williams, K. M. (2015). Effects of processing on the recovery of food allergens from a model dark chocolate matrix. *Food Chemistry, 168*, 580–587. https://doi.org/10.1016/j.foodchem.2014.07.084.
Koppelman, S. J., Jayasena, S., Luykx, D., Schepens, E., Apostolovic, D., de Jong, G. A. H., ... Maleki, S. (2016). Allergenicity attributes of different peanut market types. *Food and Chemical Toxicology, 91*, 82–90. https://doi.org/10.1016/J.FCT.2016.02.016.
Kosky, A. A., Dharmavaram, V., Ratnaswamy, G., & Manning, M. C. (2009). Multivariate analysis of the sequence dependence of asparagine deamidation rates in peptides. *Pharmaceutical Research, 26*(11), 2417–2428. https://doi.org/10.1007/s11095-009-9953-8.
Kottapalli, K. R., Payton, P., Rakwal, R., Agrawal, G. K., Shibato, J., Burow, M., & Puppala, N. (2008). Proteomics analysis of mature seed of four peanut cultivars using two-dimensional gel electrophoresis reveals distinct differential expression of storage, anti-nutritional, and allergenic proteins. *Plant Science, 175*(3), 321–329. https://doi.org/10.1016/J.PLANTSCI.2008.05.005.
Longo, G., Berti, I., Burks, A. W., Krauss, B., & Barbi, E. (2013). IgE-mediated food allergy in children. *The Lancet, 382*, 1656–1664. https://doi.org/10.1016/S0140-6736(13)60309-8.
Monaci, L., Pilolli, R., De Angelis, E., & Mamone, G. (2015). Mass spectrometry in food allergen research. *Comprehensive Analytical Chemistry, 68*, 359–393. https://doi.org/10.1016/B978-0-444-63340-8.00007-8.
Moretzsohn, M. de C., Hopkins, M. S., Mitchell, S. E., Kresovich, S., Valls, J. F. M., & Ferreira, M. E. (2004). Genetic diversity of peanut (Arachis hypogaea L.) and its wild relatives based on the analysis of hypervariable regions of the genome. *BMC Plant Biology, 4*, 11. https://doi.org/10.1186/1471-2229-4-11.
Palladino, C. (2018). Peanut allergens. *Molecular Immunology, 100*, 58–70. https://doi.org/10.1016/J.MOLIMM.2018.04.005.
Planque, M., Arnould, T., Delahaut, P., Renard, P., Dieu, M., & Gillard, N. (2019). Development of a strategy for the quantification of food allergens in several food products by mass spectrometry in a routine laboratory. *Food Chemistry, 274*, 35–45. https://doi.org/10.1016/j.foodchem.2018.08.095.
Pomés, A., Davies, J. M., Gadermaier, G., Hilger, C., Holzhauser, T., Lidholm, J., ... Goodman, R. E. (2018). WHO/IUIS Allergen Nomenclature: Providing a common language. *Molecular Immunology, 100*, 3–13. https://doi.org/10.1016/J.MOLIMM.2018.03.003.
Popping, B., & Godefroy, S. B. (2011). Allergen detection by mass spectrometry-The new way forward. *Journal of AOAC International*.
Reimer, J., Shamshurin, D., Harder, M., Yamchuk, A., Spicer, V., & Krokhin, O. V. (2011). Effect of cyclization of N-terminal glutamine and carbamidomethyl-cysteine (residues) on the chromatographic behavior of peptides in reversed-phase chromatography. *Journal of Chromatography A, 1218*(31), 5101–5107. https://doi.org/10.1016/J.CHROMA.2011.05.079.
Robinson, N. E., & Robinson, A. B. (2004). Prediction of primary structure deamidation rates of asparaginyl and glutaminyl peptides through steric and catalytic effects. *Journal of Peptide Research, 63*(5), 437–448. https://doi.org/10.1111/j.1399-3011.2004.00148.x.
Robinson, N. E., Robinson, Z. W., Robinson, B. R., Robinson, A. L., Robinson, J. A., Robinson, M. L., & Robinson, A. B. (2004). Structure-dependent nonenzymatic deamidation of glutaminyl and asparaginyl pentapeptides. *The Journal of Peptide Research, 63*(5), 426–436. https://doi.org/10.1111/j.1399-3011.2004.00151.x.
Sayers, R. L., Johnson, P. E., Marsh, J. T., Barran, P., Brown, H., & Mills, E. N. C. (2016). The effect of thermal processing on the behaviour of peanut allergen peptide targets used in multiple reaction monitoring mass spectrometry experiments. *Analyst, 141*(13), 4130–4141. https://doi.org/10.1039/C6AN00359A.
Schechter, I., & Berger, A. (1967). On the size of the active site in proteases. I Papain. *Biochemical and Biophysical Research Communications, 27*(2), 157–162. https://doi.org/10.1016/S0006-291X(67)80055-X.
Schmidt, H., Gelhaus, C., Latendorf, T., Nebendahl, M., Petersen, A., Krause, S., ... Janssen, O. (2009). 2-D DIGE analysis of the proteome of extracts from peanut variants reveals striking differences in major allergen contents. *Proteomics, 9*(13), 3507–3521. https://doi.org/10.1002/pmic.200800938.
Sicherer, S. H., & Sampson, H. A. (2014). Food allergy: Epidemiology, pathogenesis, diagnosis, and treatment. *Journal of Allergy and Clinical Immunology, 133*. https://doi.org/10.1016/j.jaci.2013.11.020.
Siepen, J. A., Keevil, E.-J., Knight, D., & Hubbard, S. J. (2007). Prediction of missed cleavage sites in tryptic peptides aids protein identification in proteomics. *Journal of Proteome Research, 6*(1), 399–408. https://doi.org/10.1021/pr060507u.
Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., ... Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Retrieved from *Molecular Systems Biology, 7*(1)http://msb.embopress.org/content/7/1/539.abstract.
Smith, J., & Hardacre, A. (2011). Development of an extruded snack product from the legume Vicia faba minor. *Procedia Food Science, 1*, 1573–1580. https://doi.org/10.1016/J.PROFOO.2011.09.233.
Van Hengel, A. J. (2007). Food allergen detection methods and the challenge to protect food-allergic consumers. *Analytical and Bioanalytical Chemistry, 389*, 111–118. https://doi.org/10.1007/s00216-007-1353-5.
Walczyk, N. E., Smith, P. M. C., Tovey, E., Wright, G. C., Fleischfresser, D. B., & Roberts, T. H. (2013). Analysis of crude protein and allergen abundance in peanuts (Arachis hypogaea cv. Walter) from three growing regions in Australia. *Journal of Agricultural and Food Chemistry, 61*(15), 3714–3725. https://doi.org/10.1021/jf305347r.
Walker, M. J., Burns, D. T., Elliott, C. T., Gowland, M. H., & Mills, E. N. C. (2016). Is food allergen analysis flawed? Health and supply chain risks and a proposed framework to address urgent analytical needs. *The Analyst, 141*(1), 24–35. https://doi.org/10.1039/c5an01457c.
Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics, 25*(9), 1189–1191. https://doi.org/10.1093/bioinformatics/btp033.
Wu, Z., Zhou, N., Xiong, F., Li, X., Yang, A., Tong, P., ... Chen, H. (2016). Allergen composition analysis and allergenicity assessment of Chinese peanut cultivars. *Food Chemistry, 196*, 459–465. https://doi.org/10.1016/j.foodchem.2015.09.070.

4.3. "Selection of universal peptide biomarkers for the detection of the allergen hazelnut in food through a comprehensive, high resolution mass spectrometric (HRMS) based approach" (Van Vlierberghe *et al*, 2020)

The strategy to identify potential peptide biomarkers was applied to hazelnut by Kaatje Van Vlierberghe at ILVO.

Allergies to tree nuts are frequent and among the latter, hazelnut continues to be globally one of the most common food allergens with an estimated prevalence of up to 4.9 % of the general population (Costa *et al*, 2016b; McWilliam *et al*, 2015). Because of their organoleptic properties, hazelnuts are not only consumed raw, but are widely used in the food industry in chocolate, confectionery, baking, ice cream and dairy products.

The developed and optimized sample preparation protocol was applied to processed hazelnut matrices: ground hazelnuts as unprocessed matrix, oven roasted ground hazelnuts as heated matrix, fermented hazelnut milk as low pH matrix, caramelised ground hazelnuts as a matrix subjected to Maillard reactions and chocolate incurred with hazelnut as fat-rich and complex matrix. Resulting peptides were analysed by HPLC-HRMS.

As for peanuts, hazelnuts from two distinct geographical origins (Turkey and Italy) were considered to evaluate the potential effect on protein content and allergen relative abundance. Each hazelnut matrix was produced and analysed in duplicate, one for each geographical origin.

Identified peptides were filtered using a set of selection criteria to ensure method specificity, sensitivity, and robustness. Ideal peptide biomarkers must be specific to hazelnut, belong to abundant proteins, be size-compatible with triple quadrupole mass spectrometry analysis, and be robust to food processing but not prone to missed cleavages by the protease(s) used or to amino acid modifications. A total of eight peptides from three major hazelnut proteins (allergens Cor a 8, Cor a 9 and Cor a 11) were identified as potential peptide biomarkers.

Compared to egg or peanut, hazelnut protein databases are rather limited (around 500 entries for hazelnut compared to 10 000 for peanut in the UniProt database). No data are currently available concerning multiple protein isoforms. The list of selected peptides should be updated once more information becomes accessible in the literature. Differences in protein abundance between the two hazelnut geographical origins were however observed for protein Cor a 9.

The eight identified potential peptide biomarkers were considered for the development of the quantitative UHPLC-MS/MS analysis method.

The research article was published in Food Chemistry Journal.

Food Chemistry 309 (2020) 125679

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

# Selection of universal peptide biomarkers for the detection of the allergen hazelnut in food trough a comprehensive, high resolution mass spectrometric (HRMS) based approach

Kaatje Van Vlierberghe[a], Maxime Gavage[b,c], Marc Dieu[c], Patsy Renard[c], Thierry Arnould[c], Nathalie Gillard[b], Katleen Coudijzer[a], Marc De Loose[a], Kris Gevaert[d,e], Christof Van Poucke[a,*]

[a] ILVO Flanders Research Institute for Agriculture, Fisheries and Food, Technology and Food Science Unit, Brusselsesteenweg 370, BE-9090 Melle, Belgium
[b] CER Groupe, Rue du Point du Jour 8, 6900, Marloie, Belgium
[c] Laboratory of Biochemistry and Cell Biology (URBC)-Namur Research Institute for Life Sciences (NARILIS), University of Namur, 61, rue de Bruxelles, 5000 Namur, Belgium
[d] VIB-UGent Center for Medical Biotechnology, Albert Baertsoenkaai 3, BE-9000 Ghent, Belgium
[e] Department of Biomolecular Medicine, Ghent University, Albert Baertsoenkaai 3, BE-9000 Ghent, Belgium

ARTICLE INFO

*Keywords:*
High-resolution mass spectrometry
Peptide biomarker selection
Food processing
Allergen
Hazelnut

ABSTRACT

The interest of using LC-MS/MS as a method for detection of allergens in food is growing. In such methods, peptides are used as biomarkers for the detection and quantification of the allergens. The selection of good biomarker peptides is of high importance to develop a specific, universal and sensitive method. Biomarkers should, for example, be robust to food processing. To evaluate robustness, test material incurred with hazelnut having undergone different food processing techniques was produced. Proteins of these materials were extracted, digested and further analyzed using HRMS. After peptide identification, selection was carried out using several criteria such as hazelnut specificity and amino acid composition. Further selection was done by comparing peptide MS intensities in the different food matrices. Only peptides showing processing robustness were retained. Eventually, eight peptides coming from three major hazelnut proteins were selected as the best biomarkers for hazelnut detection in processed foods.

## 1. Introduction

Food allergies – immunological reactions to a certain component, mostly proteins, from a certain food – form a growing health problem around the world, especially in industrialized countries (Costa, Mafra, Carrapatoso, & Oliveira, 2016). Indeed, an estimated 17 million European individuals suffer from food allergy, which is twice as much as 10 years ago (Manea, Ailenei, & Deleanu, 2016). Allergies to tree nuts, among which hazelnut, continues to be globally one of the most common food allergens with an estimated prevalence of up to 4.9% of the general population (Costa et al., 2016; McWilliam et al., 2015). Because of its organoleptic properties, hazelnuts are not only consumed raw, but are widely used in the food industry in chocolate, confectionary, baking, ice cream and dairy products. Furthermore, given their nutritional and nutraceutical benefits, there is a rise in consumption of hazelnuts in the western countries, which leads to an increase in hazelnut allergy incidence (Özenç & Özenç, 2015).

Allergic patients sensitive to hazelnuts can, after consumption, suffer from rather mild symptoms (hives, dizziness, and/or shortened breath) to very serious and even life-threatening symptoms such as an anaphylactic shock (Özenç & Özenç, 2015). As until today no cure is available to treat food allergies, patients can only protect themselves by avoiding consumption of culprit foods (Popping & Diaz-Amigo, 2017). Therefore, European legislation (Regulation 1169/2011/CE) dictates the declaration of hazelnut and other foods with known allergenic potential when deliberately used as an ingredient, regardless of the concentration or form (Popping & Diaz-Amigo, 2017). However, this is not a complete solution as unintentional presence of trace amounts of hazelnut often occurs, by cross contamination in factories, the use of contaminated ingredients or other causes (Monaci & Visconti, 2010).

* Corresponding author.
*E-mail addresses:* kaatje.vanvlierberghe@ilvo.vlaanderen.be (K. Van Vlierberghe), m.gavage@cergroupe.be (M. Gavage), marc.dieu@unamur.be (M. Dieu), patsy.renard@unamur.be (P. Renard), thierry.arnould@unamur.be (T. Arnould), n.gillard@cergroupe.be (N. Gillard), katleen.coudijzer@ilvo.vlaanderen.be (K. Coudijzer), marc.deloose@ilvo.vlaanderen.be (M. De Loose), kris.gevaert@vib-ugent.be (K. Gevaert), christof.vanpoucke@ilvo.vlaanderen.be (C. Van Poucke).

https://doi.org/10.1016/j.foodchem.2019.125679
Received 22 March 2019; Received in revised form 19 July 2019; Accepted 7 October 2019
Available online 19 October 2019

Since a very small dose of hazelnut can already elicit a reaction in allergic individuals (a few milligrams are enough to cause a severe reaction), this unintentional presence poses a high risk for patients.

To improve protection of patients, accurate, sensitive and reliable analytical methods for the detection and quantification of allergens in food are thus required (Prado et al., 2016). The detection techniques currently applied can generally be divided into two main categories: DNA-based detection (quantitative Polymerase Chain Reaction (qPCR), digital PCR), by which food allergens are detected through amplification of allergen-specific DNA or DNA fragments, and protein-based detection (e.g., Enzyme-Linked Immuno Sorbent Assay (ELISA)), in which detection is based on the binding of a specific protein of the food allergen to a specific receptor (i.e., an antibody) (Prado et al., 2016). Although these techniques are very sensitive and fast, they display some drawbacks, especially for the detection of allergens in processed foods. Indeed, most of the foods consumed in modern developed societies are processed (Sathe, Teuber, & Roux, 2005) and the applied techniques may alter a protein's structure/conformation through denaturation and/or modifications. They could also influence their extractability by formation of protein aggregates. Altering of a protein's structure can thus result in a loss of conformational epitopes, destroying recognition sites used in ELISA detection and thus generate false negative results (Prado et al., 2016). Moreover, cross reactivity of the antibody in the ELISA test with unknown substances can't be fully excluded, giving rise to false positive results (Wei, Sathe, Teuber, & Roux, 2003). DNA-based detection techniques on the other hand do not detect the elicitor (thus the protein) of the allergic reaction itself, which is thus less informative for patients (Poms, Klein, & Anklam, 2004). Indeed, food containing the allergen DNA can be harmless for patients if no allergen proteins are present, while the absence of allergen DNA cannot exclude the presence of allergen proteins.

Another technique that can be used for the detection of allergens in food and which is gaining in popularity is liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) (Pilolli, De Angelis, & Monaci, 2017). This technique is based on the detection of so-called proteotypic peptides, peptides that uniquely match to the ingredient specific protein of interest. These MS-based methods hold particular advantages including the possibility of multiplexing, high sensitivity and specificity, minimization of cross reaction and are qualitative as well as quantitative (Pilolli et al., 2017). Another advantage is the possibility to develop a detection method that is robust to processing techniques, thus universalizing the use of the method. However, in order for the method to be robust to food processing, a proper selection of proteotypic peptides is required. Such peptides have to fulfill several criteria in addition to being specific for the allergen of interest: they have to be stable during food processing and preferably should not contain any amino acids prone to (post-translational) modifications.

In this study we describe a comprehensive approach for the identification and selection of proteotypic peptides robust to food processing that can later be used as analytes for the development of an MS based method for the detection of hazelnut in food. Although several studies based on proteotypic peptides for the detection of hazelnut in food already have already been published (Ansari, Stoppacher, & Baumgartner, 2012; Bignardi et al., 2013; Costa et al., 2016; Cucu, Meulenaer, & Devreese, 2012; Heick, Fischer, Kerbach, Tamm, & Popping, 2011; Pilolli et al., 2017; Pilolli, De Angelis, & Monaci, 2018; Planque et al., 2017, 2019; Sealey-Voyksner, Zweigenbaum, & Voyksner, 2016), to the best of our knowledge, the peptides identified in these studies were not checked for food processing robustness. In the report by Pilolli and collaborators, peptide profiles generated from different food matrices having undergone several processing techniques would need to be compared. However, these matrices should be incurred (incorporation of raw ingredients before food processing), and not spiked (incorporation of extracted proteins from raw ingredients after processing of the matrix). Preferably, production of these matrices can be fully controlled, with exact knowledge of production circumstances, recipe and ingredients, and keeping cross-contamination events as low as possible. This approach will render trustworthy results, but requires an agreed science based selection of proteotypic peptides.

We tackled this problem by producing five different hazelnut-incurred food matrices. Four different processing conditions (heating, induction of the Maillard reaction, low pH environment and fatty rich environment) were chosen based on techniques that are often used in modern food industry, and material was produced using semi-industrial equipment to mimic realistic conditions of food processing as much as possible. Raw material was self-selected ensuring traceability, and these raw ingredients were used as blank (no processing) samples. These test materials were used for the generation of tryptic peptide profiles by ultra high-performance liquid chromatography coupled to high resolution mass spectrometry (UHPLC-HRMS). Proteotypic peptides were subsequently selected based on multiple selection criteria such as ingredient specificity, amino acid composition and robustness towards food processing. Eventually, we were able to identify eight peptides, originating from two hazelnut proteins – vicilin cor a 11 and legumin cor a 9, two major hazelnut allergens – suitable as good biomarkers for the detection of hazelnut in food.

## 2. Experimental

### 2.1. Material

All reagents and chemicals were of analytical grade unless stated otherwise.

Urea, tris(hydroxymethyl)aminomethane (Tris-HCl), tris(2-carboxyethyl)phosphine hydrochloride (TCEP), iodoacetamide (IAA), tetraethylammonium bicarbonate (TEAB) and NAP-10 columns (GE Healthcare) were purchased from Sigma-Aldrich (Bornem, Belgium). Trifluoroacetic acid (TFA) was obtained from Thermo Fisher Scientific (Waltham, Massachusetts, USA), and solid phase extraction cartridge (SPE) Pak C18 1 cc Vac Cartridge from Waters (Milford, Massachusetts, USA). Enzymatic digestion was performed with Trypsin Gold, Mass Spectrometry Grade from Promega (Madison, Wisconsin, USA), and Lysyl Endopeptidase, Mass Spectrometry Grade (Lys-C) from Wako Pure Chemical Industries, Ltd (Osaka, Japan). Acetonitrile, methanol absolute, 2-propanol (ULC/MS grade for all solvents), and formic acid were from Biosolve (Valkenswaard, the Netherlands). Ammonium phosphatide was kindly provided by Palsgaard A/S (Juelsminde – Denmark).

### 2.2. Production of test materials

Processed hazelnut matrices were produced in the Food Pilot unit of ILVO. All hazelnut matrices were prepared starting from raw hazelnuts from either Turkish origin (Versenoten, Alphen aan den Rijn, The Netherlands) or Italian origin (Quality Nuts BVBA, Zandhoven, Belgium). Both types of purchased hazelnut were dehulled and vacuum packed. Total nitrogen content was determined by the Kjeldahl assay. These values were multiplied with a value of 5.3 (AOAC factor for nuts) (Mariotti, Tomé, & Mirand, 2008) to calculate the protein content in both nut types. This resulted in 14.8% and 14.5% proteins for hazelnuts originating from Turkey and Italy respectively. Production of all test material was done in duplicate, once using hazelnut originating from Turkey and once using hazelnuts originating from Italy.

#### 2.2.1. Preparation of a raw (=unprocessed) hazelnut matrix

Raw whole hazelnuts were grounded using a Kenwood AT286 Glass mini chopper/mill attachment KW714229 (Kenwood, New Hamphshire, UK) for 20 s at pulsing mode using maximum speed, vacuum packed (vacuum packaging machine KN5, vc999, Missouri, USA) and stored at 4 °C in the dark.

#### 2.2.2. Preparation of a heated hazelnut matrix

Raw whole nuts were roasted in batches of 500 g in an oven (baking

oven condo CO 6.0608, Miwe, Germany) at 180 °C for 18 min. Nuts were spread in one layer on a baking plate, and each 5 min, nuts were stirred with a spoon to ensure homogenous roasting. Roasted nuts were left to cool down at room temperature, mixed into a nut powder, vacuum packed and stored at 4 °C.

*2.2.3. Preparation of a hazelnut matrix subjected to the Maillard reaction (= caramelized nuts)*

250 g raw whole nuts were chopped (Blixer 4VV -, 4,5 Liter, 1100 W, Variable Velocity: 300–3.500 TPM, Robot Coupe, Vincennes Cedex, France) for 3 times 5 s at maximum speed. An amount corresponding to 200 g lactose and 100 mL reverse osmose water were mixed into a smooth paste. The chopped nuts were subsequently stirred into the lactose paste until nut parts were all covered with the lactose substance. This nut-lactose mixture was spread over an oven plate and placed in a convectional pre-heated (160 °C) oven for 15 min. Every 5 min, the mixture on the plate was stirred with a spoon to obtain equal caramelization. The plate was left at room temperature to cool down overnight, and subsequently mixed into a fine power (Kenwood Mini Chopper Mill Attachment AT320B). The caramelized nut powder was divided into portions of 50 g, vacuum packed and stored at 4 °C in the dark.

*2.2.4. Preparation of a low pH hazelnut matrix (= fermented hazelnut milk)*

This matrix was produced by fermenting nut milk using a "set yoghurt" procedure. Nut milk was produced by soaking the nuts in reverse osmose water overnight, followed by mixing the substance (Blixer 4VV -, 4.5 Liter, 1100 W, Variable Velocity: 300–3.500 TPM, Robot Coupe, Vincennes Cedex, France) for 8 min at maximum speed upon which a milky substance was reached. Nut milk was subsequently sieved (Retsch Test Sieve, 250 μm pore size, Retsch, Aartselaar, Belgium). Sieved milk was then transferred to a sterile Erlenmeyer of 1 L, and a bacterial culture (nu-trish® BY-Mild – containing the following bacteria: *Bifidobacterium* species, *Lactobacillus delbrueckii* subsp. *Bulgaricus* and *Streptococcus thermophiles*, CHR Hansen, Hørsholm, Denemarken) was added (4 g of frozen culture for 600 mL of nut milk). Erlenmeyer recipients were aseptically sealed and incubated in a warm water bath (42 °C). Every 30 min, a sub-sample of the yoghurt was checked for pH until a pH of 4.2 was reached. The yoghurt was subsequently divided into 20 g portions and stored at 4 °C in the dark.

*2.2.5. Preparation of a fatty rich hazelnut matrix (= chocolate incurred with hazelnut)*

A nut powder was produced by first mixing raw nuts (Kenwood Mini Chopper Mill Attachment AT320B, pulsing speed, 3 min, Kenwood, New Hamphshire, UK), resulting in a slushy substance. Nut mash was then defatted by pressing it between extra thick kitchen towels, resulting in a dry nut powder. The defatting step was done to obtain a very fine grain size to ensure a homogenous chocolate after mixing it in. Nut mash and powder was weighed before and after defatting, and loss in mass was documented to determine how much weight was lost during defatting. This fine, dry nut powder was then sieved through a pore size of 500 μm. Secondly, chocolate was prepared: first, a "conche" was made from chocolate refiner flakes (90%) and cacao butter (10%). 450 g refiner flakes (Callebaut, Lebbeke, Belgium) and 50 g cacao butter (Callebaut) were therefore warmed in a water bath in separate bowls until a temperature of 40–45 °C was reached. The temperature was controlled with a food thermometer. Then, the melted butter was slowly added to the melted refiner flakes, and the mix was then stirred for 3 min before addition of 1.2% ammonium phosphatide. The mixture was then stirred for an extra 3 min. Finally, a chocolate incurred with hazelnut (40,207 ppm total hazelnut proteins in matrix for hazelnut origin Italy and 41,046 ppm for hazelnut origin Turkey) was produced by adding 99.996 g of defatted hazelnut powder (containing 24.63 g proteins for hazelnut origin Turkey and 24.12 g proteins for hazelnut origin Italy) to the liquid chocolate and stirring again for 3 min while keeping the temperature of the chocolate at 42 °C. Ppm levels are mentioned as mg total hazelnut proteins/kg final chocolate product. Liquid chocolate was subsequently poured into chocolate molds, covered with plastic foil and left to cool down. Chocolate was kept overnight at 4 °C to cool further, set and solidify. The next day, chocolate bars were rasped into chocolate flakes. Chocolate flakes were kept in aliquots of 50 g and stored at 4 °C in the dark. A proper blank chocolate was prepared following the same procedure, without adding the hazelnut powder.

*2.3. Sample preparation*

Sample preparation was similar to the protocol described in our previous study (Gavage et al., 2019). Briefly, proteins were extracted from food matrices, purified and proteolytically digested. From each matrix of both hazelnut origins, three biological repeats were prepared and analyzed by LC-HRMS.

*2.3.1. Protein extraction*

An amount of 1 g of matrix material was extracted with 10 mL extraction buffer (2 M urea, 200 mM Tris-HCl, pH 9.2) by shaking at 20 °C for 30 min at 300 rpm, followed by 15 min sonication on ice at 100% ultrasound power. The samples were centrifuged at 4600*g* for 10 min at 4 °C after which the middle liquid phase was transferred to a new tube. The protein concentration of the extract was subsequently determined with A280 ultra-violet absorbance protein assay (NanoPhotometer® N60, Westburg, Leusden, Nederland) (Desjardins, Hansen, & Allen, 2009).

*2.3.2. Protein reduction, alkylation and purification*

Samples were diluted with extraction buffer to a protein concentration of 4 mg of proteins per mL. Next, proteins were reduced and alkylated (for 15 min at 37 °C in the dark) with TCEP-HCl (15 mM final concentration) and IAA (30 mM final concentration) under agitation at 500 rpm. The samples were desalted and buffer-exchanged by gel filtration chromatography (NAP-10 columns, GE Healthcare, United Kingdom). Proteins were eluted in 1.5 mL 6 M urea in 50 mM TEAB. The protein concentration was measured with the A280 ultra-violet absorbance protein assay.

*2.3.3. Enzymatic digestion*

An amount corresponding to 100 μg of proteins was transferred to a new micro centrifuge tube and diluted with 6 M urea in 50 mM TEAB to a final volume of 50 μL before addition of 2 μg of Lys-C, and digestion was performed for 2 h at 37 °C under 300 rpm agitation. The samples were diluted with 50 mM TEAB to a final volume of 300 μL, and 1 μg of trypsin was then added. Tryptic digestion was subsequently performed for 16 h digestion at 37 °C under 300 rpm agitation. Digestion was stopped by adding TFA until a pH of ≤ 3 was reached.

*2.3.4. Peptide purification*

Peptides were purified with C18 SPE cartridges. The cartridges were first washed with 1 mL acetonitrile/water (50/50, v/v) and equilibrated with 1 mL wash solvent (acetonitrile/water/TFA, 2/97.9/0.1, v/v/v). Samples were diluted 10 times with water, loaded onto the cartridges and washed with 1 mL of wash solvent. Peptides were eluted with 2 mL acetonitrile/water (80/20, v/v) and subsequently evaporated under a nitrogen flow at 40 °C. Peptide pellets were dissolved in 200 μL 0.1% formic acid and filtered (centrifugal filter unit 0.22 μm, Merck-Millipore, Carrigtwohill, Ireland).

*2.4. Mass spectrometry*

Solubilized peptides from a digest of 100 μg proteins were analyzed by high performance liquid chromatography coupled to high resolution

mass spectrometry using a Waters Acquity UPLC® (Waters, Milford, MA VS) coupled to a SYNAPT G2-S High Definition Mass Spectrometer (TOF) (Waters). Tryptic peptides were chromatographically separated on a ACQUITY UPLC Peptide BEH C18 column (300 Å, 1.7 µm, 2.1 mm X 150 mm) (Waters) protected by an ACQUITY UPLC Peptide BEH C18 VanGuard Precolumn (300 Å, 1.7 µm, 2.1 mm X 5 mm) (Waters). Mobile phase A was 0.1% formic acid (solvent A) and mobile phase B was 0.1% formic acid in acetonitrile (solvent B). The solvent gradient was set as follows: initial: 99% solvent A; 60 min: 60% solvent A; 61 min: 15% solvent A; 65 min: 99% solvent A. The injection volume was set at 5 µL, the flow rate at 0.200 mL/min. Peptides were ionized in positive electrospray (ESI + ) with the following instrument settings: capillary voltage 0.80 kV, sample cone voltage: 80 V, source temperature 120 °C and desolvation temperature: 400 °C. Data acquisition was done in continuous full scan mode in $MS^E$ mode (data independent acquisition) and the *m/z* range was 50–2,000. Fragmentation was done with collision induced dissociation, with a low trap collision energy for precursor ions of 4 eV and high collision energy ramping from 20 to 40 eV. Recording was done in resolution mode (20 k FWHM). Prior to analysis, the HRMS was calibrated with sodium formate. Leucine enckephalin (200 pg/µL, mass 556.2771 Da, Waters) was used as lock mass and was acquired (scan time of 0.2 s) with an interval of 30 s, but not yet applied for mass correction. Mass correction using lock mass acquisitions is intrinsically imbedded in the ProteinLynx data analysis algorithm.

### 2.5. Data analysis

Raw HRMS data were analyzed with ProteinLynx Global Server™ (PLGS, Waters). Raw data were first automatically mass corrected using the lock mass recordings. Identified peaks were searched against a database containing the hazelnut proteome stored in Uniprot.org (*Corylus avellana*, 474 entries), added with the 192 reviewed human keratin entries from Uniprot.org, the cationic bovine trypsin sequence, the Lysyl endopeptidase sequence from *Lysobacter enzymogenes* and the possible protein contaminants from cacao and bacteria from Nu-trish® BY-Mild (*Bifidobacterium animalis* subsp. *lactis*, Lactobacillus *delbrueckii* subsp. *bulgaricus* and *Streptococcus thermophilus*). Search parameters were as follows: trypsin was chosen as the primary digestion enzyme and carbamidomethylation of cysteine residues was set as a fixed modification. Oxidation of methionine residues, conversion of *N*-terminal glutamine to pyroglutamate, and deamidation of asparagine and glutamine were set as variable modifications and 3 missed cleavages were allowed.

## 3. Results and discussion

### 3.1. Test material and hazelnut origin

Our main goal was to identify peptides originating from hazelnut proteins that are robust towards food processing and could thus serve as universal/excellent biomarkers for the detection of hazelnut in food. Therefore, we extracted peptide profiles from different hazelnut-containing food matrices that underwent various processing techniques. Through comparison of these profiles, we expected to identify peptides robust to processing. To obtain these peptide profiles, food matrices needed first to be selected. However, when purchasing such materials from supermarkets, one can never be sure about the raw material that was used, recipes, food processing conditions, and/or putative cross contamination. Therefore, test material was produced in-house using modern food processing techniques and equipment. The four well-controlled processing conditions were heating of the ingredient, induction of the Maillard reaction, incurring hazelnut in a low pH environment or in a fatty rich environment. This generated the following test materials: roasted hazelnuts, caramelized hazelnuts, fermented hazelnut milk and chocolate incurred with hazelnut, and raw hazelnuts as a non-processed matrix.

A universal detection method for hazelnut implicates that this method should be able to detect hazelnut irrespective of the processing technique the food has undergone, but also irrespective of the origin of hazelnut. However, numerous hazelnut cultivars exist and different cultivars may have different protein profiles and, subsequently, different peptide profiles. Indeed, Garino and collaborators have recently shown that gene transcript levels of known allergenic proteins Cor a 8 and Cor a 11 in hazelnut differ between cultivars and also between year of harvest (Garino et al., 2013). To counter this issue at least partially, all test matrices were produced in “duplo” using hazelnut from two different origins. Based on worldwide production, 70% of all hazelnuts come from Turkey and the second largest producer, counting for 12% of worldwide production, is Italy (Özenç & Özenç, 2015). Hazelnuts from these two origins were therefore used for the production of test material.

### 3.2. Protein selection

The first step in selecting suitable biomarker peptides is the *in silico* selection in the UniProt database of the proteins from which peptides can be derived. In this study, we initially focused on all known hazelnut proteins, not only on proteins known to cause allergic reactions as this could lead to excluding potentially good biomarkers from high abundant proteins. The full UniProt database for *Corylus avellana* containing 474 entries was therefore used. This is a relatively low amount of entries, considering the 28,255 transcript contigs – of which 93.3% encode proteins – in the hazelnut genome (Rowley et al., 2012). However, hazelnut kernels consist of about 15% proteins, of which 87% are seed storage proteins (Bonvehí, 1995). Of these seed storage proteins in hazelnut, the globulins Cor a 9 (11S globulin, primary accession number Q8W1C2) and Cor a 11 (7S globulin, Q8S4P9), and the albumin Cor a 14 (2S albumin, D0PWG2) are the major ones. Given their high abundance, these proteins were expected to provide the most abundant peptides. Indeed, the high abundance of these proteins was evident after data analysis as identified peptides mainly came from these three proteins in all conditions tested and in hazelnuts of both origins (Table S1). A fourth protein that was also always identified is the lipid transfer protein Cor a 8 (Q9ATH2). Therefore, while present in a lower abundance in the hazelnut kernel (Bonvehí, 1995), this protein was also kept for further peptide selection. All four proteins are known as allergenic proteins and have been used for developing LC-MS/MS methods for the detection of hazelnut in food (Ansari et al., 2012; Bignardi et al., 2013; Costa et al., 2016; Cucu et al., 2012; Heick et al., 2011; Pilolli et al., 2017; Sealey-Voyksner et al., 2016).

### 3.3. Peptide selection

For LC-MS/MS analysis, peptides are required. Usually, trypsin is the protease of choice because it cuts peptide bonds C-terminal to lysine and arginine with high specificity (Hustoft et al., 2012). As such, trypsin also yields basic peptides of a suitable length (on average 10–20 amino acids) and mass for LC-MS/MS analyses. For each of the hazelnut matrices that underwent the different processing techniques, tryptic peptide profiles were generated upon analysis of the HRMS data. Peptides were considered identified if they were detected in all three biological repeats and in extracts of both origins with high confidentiality (score 2 in ProteinLynx, as analysed with ProteinLynx GlobalSERVER v3.0.3 (Waters Corporation). Parallel to this, theoretical tryptic peptide profiles for the four identified proteins were generated by *in silico* tryptic digestion using PeptideMass (Wilkins et al., 1997), avoiding any missed cleavages and only considering peptides longer than 6 amino acids to ensure specificity. HRMS results from raw hazelnut extracts were compared with this list of peptides. We were able to detect 21 out of the possible 24 peptides for Cor a 9, 19 out of 24 for Cor a 11, 5 out of 7 for Cor a 14 and 7 out of 9 potential peptides for Cor a 8 (Table S1).

*3.3.1. Specificity to hazelnut*

As the peptide biomarkers for detection of hazelnut in food should only mark the presence of hazelnut proteins, the specificity for hazelnut proteins of each peptide was first checked by performing a Basic Local Alignment Search Tool (BLAST) analysis with the peptide sequence against the full UniProt database. Peptides were considered to be specific to hazelnut only if they showed full identity to only *Corylus avellana,* and not to any other organism. Implementing this criterion resulted in the exclusion of 2 out of the 21 identified peptides for Cor a 9, among which the peptide $^{364}$Trp-Arg$^{371}$, used as a biomarker for detection of hazelnut in previous studies (Ansari et al., 2012; Costa, Ansari, Mafra, Oliveira, & Baumgartner, 2014; Gu et al., 2018), 2 out of 19 identified peptides from Cor a 11, and 2 peptides out of the 7 identified peptides from Cor a 8. All identified peptides from Cor a 14 were retained (Table S1).

*3.3.2. Amino acid composition*

In a second selection step, peptides containing certain amino acids such as methionine, cysteine and *N*-terminal glutamine were excluded. Indeed, we hypothesized that quantification of hazelnut in food products should be reflected by the intensity of the peptide biomarkers used. Thus, these peptides should be ideally present in one form with thus one particular mass. As amino acid modifications alter the mass of peptides, these may result in inaccurate quantification as such modifications can alter both the retention time of the affected peptide as well as precursor and fragment ion masses (Fig. 1B). Thus, we attempted to avoid peptides that were prone to modifications. For example, methionines in the peptide $^{105}$Gln-Arg$^{115}$ from Cor a 14 can have multiple oxidation states (each with a mass shift of + 15.995 Da). Indeed, we detected multiple forms of this peptide in all matrices (Fig. 1A). Moreover, these results show that the intensities of the peptides with such different oxidation states vary depending on the processing techniques applied, implying that quantification cannot be only based on the intensity of the non-oxidized peptide. Thus, peptides containing methionine were therefore excluded.

Along this line, we also decided to exclude cysteine-containing peptides. Although cysteines are modified by carbamidomethylation, there is a possibility that, depending on the food matrix, some cysteines are not fully reduced and modified. In addition, peptides starting with glutamine were also excluded due to the possibility of this *N*-terminal glutamine to convert to pyroglutamic acid (mass shift of −17.027 Da) (Neta, Pu, Kilpatrick, Yang, & Stein, 2007). In fact, pyroglutamic acid was also shown to hamper peptide fragmentation, which might negatively influence the overall performance of the LC-MS/MS detection method (Godugu, Neta, Simón-Manso, & Stein, 2010). The peptide $^{418}$Gln-Lys$^{432}$ for example, cited in several studies as a biomarker (Ansari et al., 2012; Costa et al., 2014; Heick et al., 2011; Pilolli et al., 2018; Planque et al., 2017, 2019), was therefore excluded. Furthermore, we selected against KP or RP motifs in peptides as trypsin might cleave such motifs, albeit with a lower efficiency. Peptides carrying other types of missed cleavages (KK, KR, RR or RK at the beginning or end of their sequence) were also excluded. The sequence motif NG was also avoided, based on the findings that the asparagine (N) in such a motif is susceptible to deamination (Jia & Sun, 2017). Finally, peptides containing the *N*-terminus of a protein were also excluded as those peptides often comprise signal peptides, which are cleaved off once the protein has reached its proper cellular location. This results in both the presence and the absence of these sequences in the extracted proteins, again leading to biased quantification. Also, *N*-terminal acetylation (Ree, Varland, & Arnesen, 2018) and excision of the *N*-terminal initiator methionine (iMet) residue from nascent peptide chains (Jonckheere, Fijałkowska, & Van Damme, 2018) can influence the detection of these peptides, thus making them unsuitable as biomarkers.

Applying these exclusion criteria withheld 10 peptides for Cor a 11, 7 for Cor a 9, 0 for Cor a 14 and 2 for Cor a 8 (Table S1). Hence, these criteria exclude a large number of identified peptides, resulting in the complete loss of Cor a 14, as well as some peptides with high intensities that could, if withheld, strengthen the power of the detection method. Indeed, an LC-MS/MS diagnostic method is preferably based on multiple peptides from multiple proteins of the ingredient. A solution to keep more proteins and peptides for final selection is to chemically induce protein modifications during sample preparation. For example, one could use hydrogen peroxide to oxidize methionines. However, such chemical modifications are only useful if they completely and specifically convert the targeted amino acids.

*3.3.3. Selection based on protein specific properties and observations*

Additional selection was done based on protein-specific properties and experimental observations. Each protein was surveyed for sequence-specific knowledge, focusing on sequence conflicts in the database, known isoforms and known natural variants. To ensure the robustness of the method, there should be no variations present in the selected peptides. For example, as Cor a 14 has two known isoforms ($^{48}$Arg ↔ $^{48}$Ser) (Pfeifer et al., 2015), the tryptic peptide $^{44}$Gly-Arg$^{53}$ was excluded. As another example, Cor a 11 has two known potential glycosylation sites (Asn$^{85}$ and Asn$^{301}$). Due to the uncertainty of these asparagine residues being glycosylated, peptides containing them were excluded (Lauer et al., 2004). Information directly extracted from the experimental data was also used for peptide selection. For example, peptides present in a sequence stretch prone to missed cleavages were excluded. The sequence $^{395}$Gly-Arg$^{419}$ in Cor a 11 is expected to yield only two fully tryptic peptides larger than 6 amino acids: $^{395}$Gly-Arg$^{403}$ and $^{407}$Glu-Arg$^{415}$. However, after HRMS analysis, various peptides were observed containing missed cleavages (Fig. 2). Again, quantification of hazelnut in food cannot be directly based on the intensity of the two fully tryptic peptides, as their true intensities are the sum of their intensities and those of all peptides that embed them. Such peptides are therefore also excluded. For example, the peptides $^{464}$Ala-Arg$^{478}$ (Ansari et al., 2012; Costa et al., 2014; Heick et al., 2011; Pilolli et al., 2017, 2018; Planque et al., 2017, 2019) and $^{407}$Glu-Arg$^{415}$ (Ansari et al., 2012; Costa et al., 2014) although often used as biomarkers in several studies, lie within a sequence stretch prone to missed cleavages. Since there is a known potential isoform (although partially sequenced, and not yet registered) for Cor a 9 (Nitride et al., 2013), the peptides $^{351}$Ile-Arg$^{363}$, $^{446}$Thr-Arg$^{458}$ and $^{464}$Ala-Arg$^{478}$ should also be excluded from selection since they lie in a non-conserved sequence region. These three peptides are also often cited as biomarker peptides for the detection of hazelnut (Ansari et al., 2012; Costa et al., 2016; Heick et al., 2011; Pilolli et al., 2017, 2018; Planque et al., 2017, 2019; Sealey-Voyksner et al., 2016). Applying these criteria led to the exclusion of another 2 peptides from Cor a 11 and 3 peptides from Cor a 9 (Table S1). Applying these selection steps eventually resulted in 2 potential good biomarkers for Cor a 8, 4 for Cor a 9 and 8 for Cor a 11.

*3.3.4. Selection based on food process robustness*

Further selection of peptides was done by comparing peptide intensities obtained in the five food matrices, and thus, controlling the peptide robustness towards the different food processing techniques. As already emphasized, peptide biomarkers that could be used for detection of hazelnut in food have to be robust to food processing, and thus, the amino acid sequence of this peptide should remain stable during the various processing steps. Also, the intensity of the peptides should be comparable under different circumstances in order to consider the peptides as good biomarkers. Sensitivity to modifications introduced by food processing can vary depending on the protein and on the peptides contained in the protein.

*3.3.4.1. Robustness towards roasting and caramelization.* In case of roasted and caramelized nuts, peptide intensities can be directly compared as no proteins other than hazelnut proteins are present in the matrix. From each extract, 100 µg of proteins were digested and peptide intensities were compared to the intensities observed for

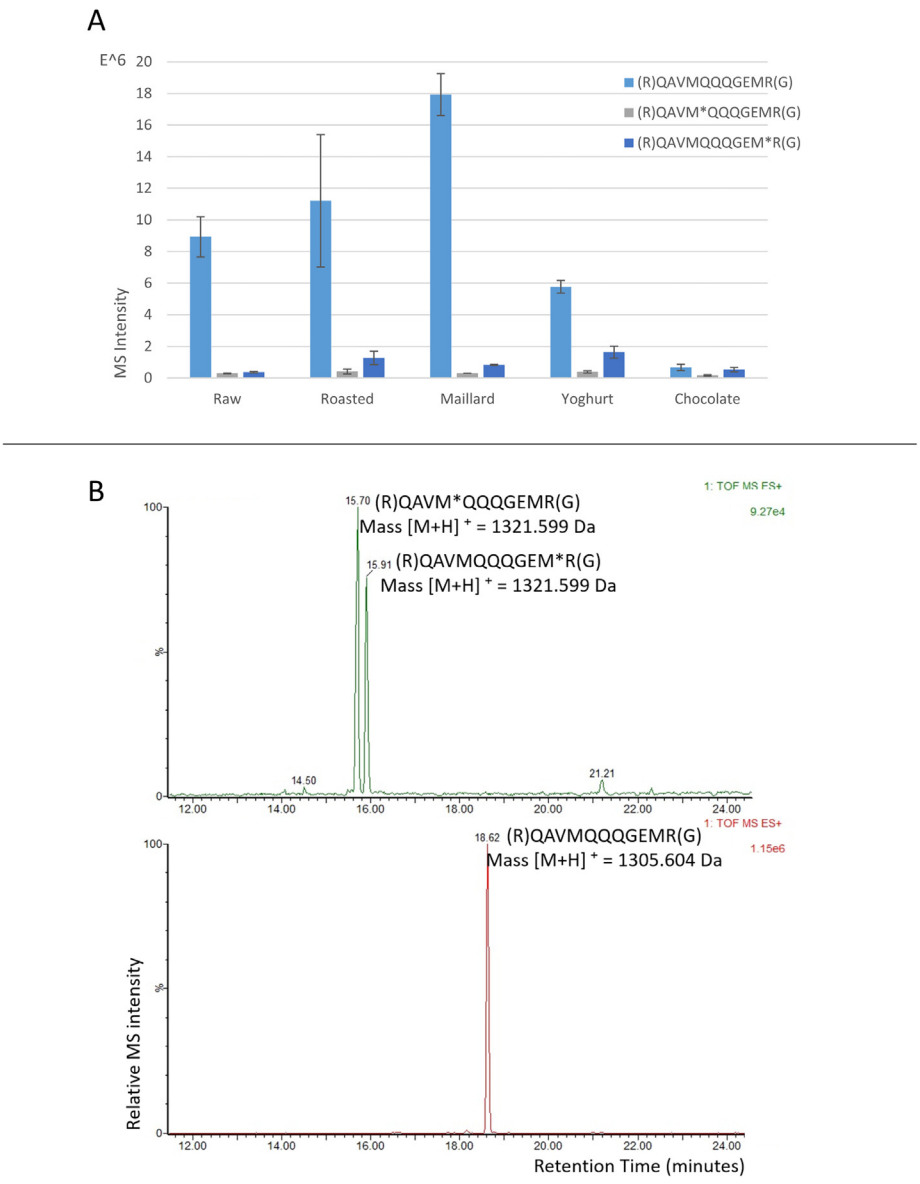



**Fig. 1.** Methionine oxidation variation. A) MS intensities of the peptide (R)QAVMQQQGEMR(G) from Cor a 14 (D0PWG2) showing different methionine oxidation states after different processing techniques. MS signal intensities are depicted for non-processed hazelnut, roasted hazelnuts, hazelnuts that have undergone induction of Maillard reaction, hazelnuts incurred in a low pH environment (yoghurt) and hazelnuts incurred in a fatty rich environment (chocolate). In each matrix, the intensities of three forms of the peptide are given. In light blue, the native (non-oxidized) peptide, in grey the peptide where the first methionine oxidized (M*) and in dark blue the peptide where the second methionine is oxidized (M*). Results are expressed as the means of the peptide MS signal intensity ± 1 SD for 3 biological replicates. B) Selected ion chromatograms of the three variants (oxidized and reduced) of the peptide (R)QAVMQQQGEMR(G) identified in roasted hazelnuts, which illustrate the difference in retention time of the different forms. Masses are shown for the single charged peptide ion $[M + H]^+$, peak intensities are shown as relative intensity with 100% = peak area of highest peak for each given mass ($9.27e^4$ for $MH^+$ = 1321.599 Da, $1.15e^6$ for $MH^+$ = 1305.604 Da). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

peptides generated from the raw sample. These peptide intensities were thus used to select biomarkers robust to processing. Peptide intensities from chocolate and yogurt carry some uncertainty as will be explained later, and will therefore only serve as an indication to the processing robustness of the selected biomarkers.

At the end of the selection steps, only 2 ($^{81}$Gly-Lys$^{96}$ and $^{52}$Ala-Arg$^{63}$) out of the 8 peptides initially identified for Cor a 8 were retained. The peptide $^{81}$Gly-Lys$^{96}$ proved to be sensitive towards

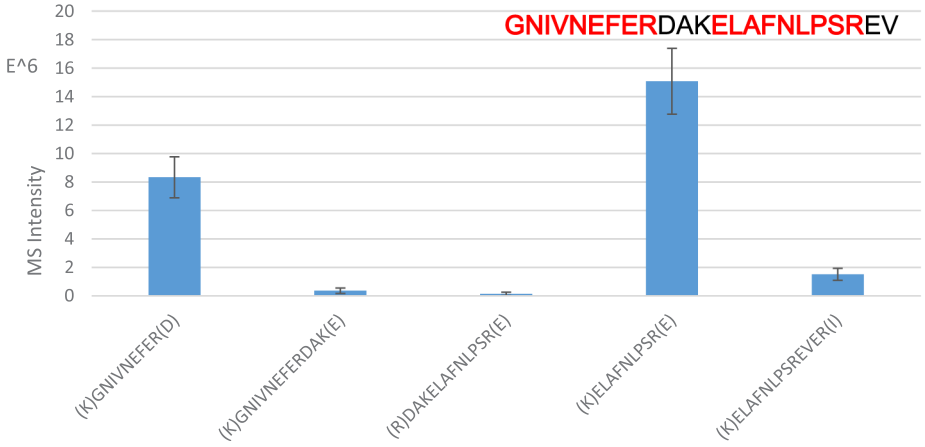



**Fig. 2.** Comparison of MS intensities of different peptides in a missed cleavage sensitive sequence stretch GNIVNEFERDAKELAFNLPSREVER from Cor a 11 (Q8S4P9). Tryptic peptides without missed cleavage are depicted in red (GNIVEFER and ELAFNLPSR). MS signal intensities are shown for raw hazelnut and results are expressed as the means of the peptide MS signal intensity ± 1 SD for 3 biological replicates. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

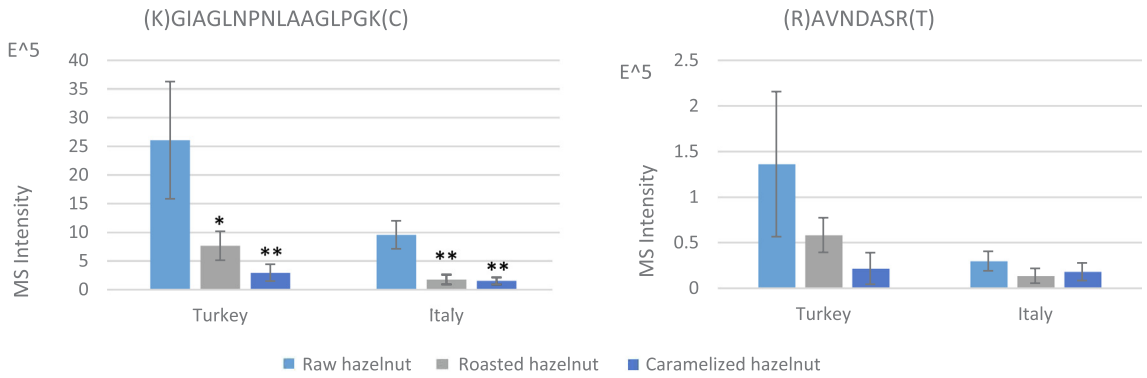



**Fig. 3.** Comparison of MS intensities of tryptic peptides that were retained after several selection criteria from Cor a 8 (Q9ATH2) in three different test materials (non-processed hazelnut (Raw hazelnut), heated hazelnuts (Roasted hazelnut) and hazelnuts that have undergone induction of Maillard reaction (Caramelized hazelnut). MS signal intensities are expressed as the means of the peptide MS signal intensity ± 1 SD for 3 biological replicates. Significance levels were calculated by Student's *t*-test, comparing values from raw hazelnut to processed hazelnut. (* = $p < 0.05$; ** = $p < 0.01$, ** = $p < 0.001$).

processing, showing significantly lower peptide intensity in both processed matrices when compared to the intensity in raw hazelnut, and this in nuts of both origins (Fig. 3). The peptide $^{52}$Ala-Arg$^{63}$, although showing an apparent decrease in intensity in processed matrices when compared to the raw matrix in hazelnut from Turkey, this decrease was not significant. No significant difference in intensity between processed and raw hazelnut matrices in hazelnuts from Italy was observed.

Cor a 11 holds both food processing sensitive as well as robust peptides. Three peptides were robust towards roasting and the Maillard reaction: $^{92}$Leu-Arg$^{100}$, $^{175}$Ile-Arg$^{201}$ and $^{241}$Ala-Arg$^{251}$. For these peptides, no significant difference in intensity was observed between the raw, roasted and caramelized hazelnut matrices (Fig. 4). Of these, $^{92}$Leu-Arg$^{100}$ and $^{175}$Ile-Arg$^{201}$ are the most intense peptides, and considering the intensities in the chocolate and fermented nut milk matrix, form the most promising biomarker candidates for this protein. The same trend was observed in samples from both nut origins. Interestingly, the peptide $^{202}$Ala-Lys$^{213}$, often cited as a biomarker (Ansari et al., 2012; Costa et al., 2014), was excluded from selection because of its sensitivity towards roasting and caramelizing hazelnut.

Selected peptides for Cor a 9 generally showed good robustness towards both roasting and caramelization. No negative significant differences in intensity were observed in peptide intensity between the three matrices (Fig. 5). Some peptides display a higher peak intensity in processed matrices when compared to the raw hazelnut, but this is not a problem for the peptide to serve as a biomarker. Cor a 9 shows overall the highest robustness across all processing techniques when compared to the other proteins. This is in line with the intrinsic properties of this protein being a legumine, and therefore very stable against thermal processes. The secondary structures of these proteins largely remain unchanged and have only minor modifications, even after heating at 94 °C (Costa et al., 2016). This while Cor a 11, a vicilin is less thermostable, showing already structural changes at 70 °C, and is known to form gellish aggregates at high temperatures, making it harder to extract the proteins and thus more difficult to detect the peptides (Sathe et al., 2005).

*3.3.4.2. Robustness towards mixing into a fatty rich environment (chocolate matrix) and in a low pH background (fermented nut milk).* When yoghurt and chocolate samples are considered, intrinsic to the recipe, non-hazelnut proteins are also present in the matrix. These proteins are co-extracted during sample preparation, digested, and their peptides are also analyzed by HRMS. Thus, in these matrices, only a fraction of the injected peptides comprises hazelnut peptides. This fraction can be theoretically calculated in the case of chocolate. The produced hazelnut incurred chocolate contains 491,099 ppm total chocolate proteins per kg chocolate, 40,708 ppm or 39,860 ppm total hazelnut proteins for Turkey or Italy hazelnut origin, respectively. Purified proteins extracted from this matrix thus comprises 55.6% cacao proteins and 44.4% hazelnut proteins, assuming cacao proteins and hazelnut proteins are extracted with the same efficiency. Peptide intensities of hazelnut peptides are therefore expected to be around 55.6% lower when compared to the peptide intensities from the same amount of extracted proteins from raw hazelnut. However, when comparing intensities from raw extracts and chocolate extracts, none of the identified peptides showed intensities in this range. On the contrary, intensities were only a fourth of the intensities in the raw sample, or even lower (Fig. S1a). An explanation could be that cacao proteins are more efficiently extracted from the chocolate matrix, enlarging the ratio cacao/hazelnut proteins in the extract. Excluding the possibility that the lower intensities are due to wrong estimation of protein content in the chocolate (indeed, polyphenols residues can influence the nanodrop results), or that protein loss during the desalting step (using NAP-10 column) is the reason for this lower abundancy, a different extraction protocol, adapted from Planque and collaborators (Planque et al., 2016) was used, where no correction for protein abundance or a desalting step is done. This resulted again in lower intensities than expected, indicating that it might be hard to detect these peptides in fatty rich environments (Fig. S1b). For the yoghurt, the theoretical ratio of hazelnut proteins to bacterial proteins is not so straightforward to calculate, due to the uncertainty on the bacterial growth rate in the fermented nut milk, and thus the uncertainty on the bacterial protein content. As could be expected, lower peptide intensities compared to the intensities from raw hazelnut were identified, but a comparison of intensities from these two matrices can thus indicate robustness to these food processing techniques, but these results should be taken with caution and are only indicative.

*3.3.5. Selection based on robustness to origin variation*

Finally, all peptides were checked on similarity of performance in the two hazelnut origins. Robustness towards genetic variations and cultivars strengthens the power of the detection method. Although in general, the same peptides showing the most promising results to serve as good biomarkers were identified in both types of hazelnut, there were some differences in peptide intensity detection. Interestingly, peptides from Cor a 9 usually display a higher intensity in Turkish hazelnuts when compared to Italian hazelnuts (Fig. S2). This trend was visible in all matrices but was more pronounced in roasted hazelnuts. This implies that the protein Cor a 9 was relatively more abundant in the Turkish hazelnuts. Also striking was that peptide intensities from roasted Turkish hazelnut were for the three proteins overall significantly higher compared to peptide intensities from roasted Italian hazelnuts, even when there was no significant difference in intensities from raw nuts (Fig. S2). Peptides from Turkish hazelnuts thus seemed to be more robust towards roasting of the nuts. This was not apparent for

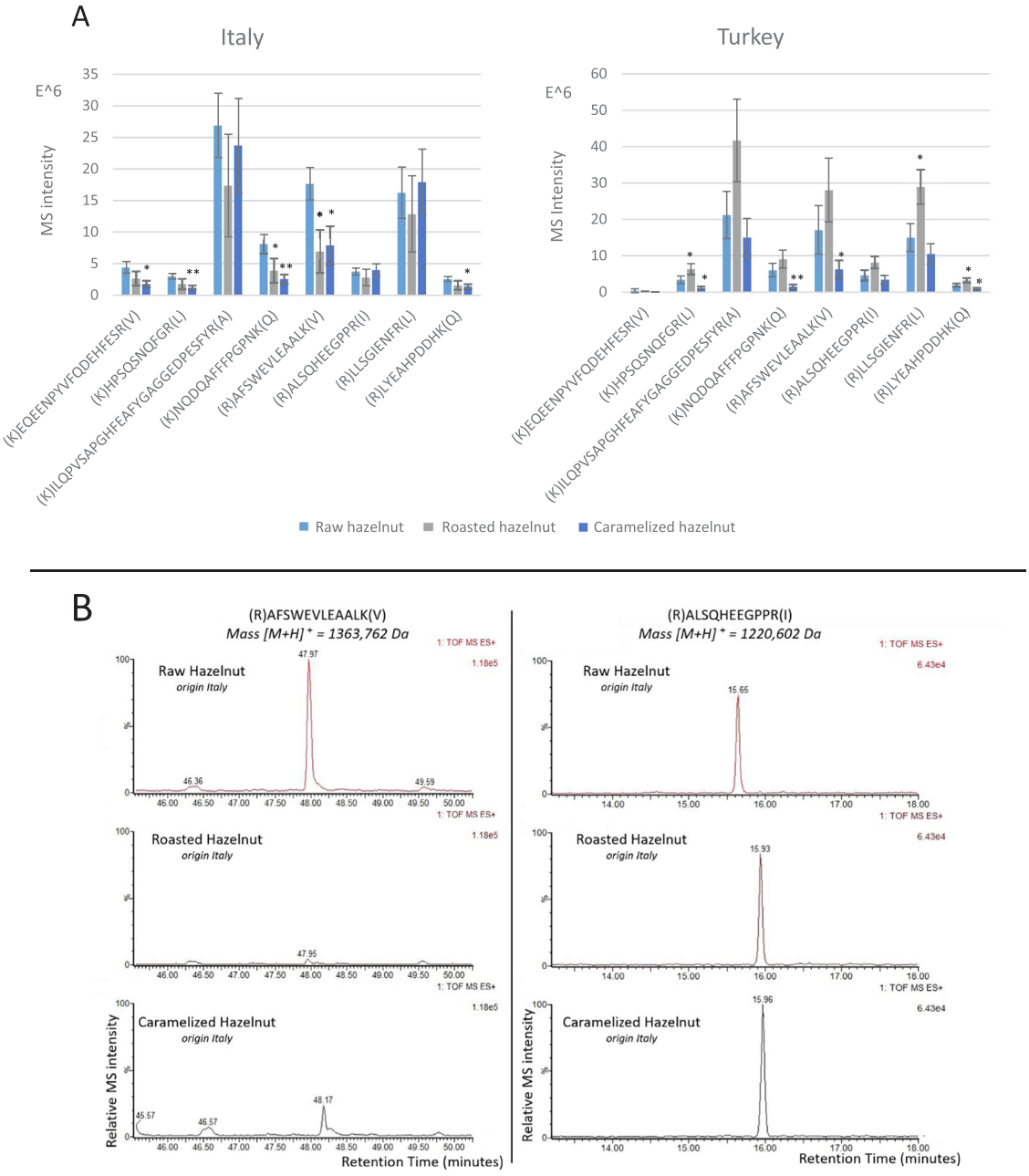



**Fig. 4.** Comparison of tryptic peptides that were retained after several selection criteria from Cor a 11 (Q8S4P9) in three different test materials (non-processed hazelnut (Raw hazelnut), heated hazelnuts (Roasted hazelnut) and hazelnuts that have undergone induction of Maillard reaction (Caramelized hazelnut). A) Comparison of MS signal intensities. Intensities are expressed as the means of the peptide MS signal intensity ± 1 SD for 3 biological replicates. Significance levels were calculated by Student's *t*-test, comparing values from raw hazelnut to processed hazelnut (* = $p < 0.05$; ** = $p < 0.01$, ** = $p < 0.001$). B) Exemplary selected ion chromatograms of a peptide showing significant difference in intensity in the different test materials ((R)AFSWEVLEAALK(V)), and a peptide showing no significant difference in intensity in the different test materials ((R)ALSQHEEGPPR(I)). Masses are shown for the single charged peptide ion $[M + H]^+$, peak intensities are shown as relative intensity with 100% = peak area of highest peak for the given mass in the three test materials ($1.18e^5$ for (R)AFSWEVLEAALK(V), $6.43e^4$ for (R)ALSQHEEGPPR(I)).

caramelization of nuts.

## 4. Conclusions

Based on the above selection steps we were able to select 8 peptides from 3 major hazelnut proteins that are suitable to function as biomarkers for detection of hazelnut in food, among which, 5 have not been published yet (Table 1).

These peptides were selected to perform good as universally applicable biomarkers, robust against multiple processing techniques and matrix effects, and insensitive to post-translational modification of amino acids occurring during production of food or during sample preparation for analysis. Of these 8 peptides, some seem to be more preferred than others. Indeed, in general, the three selected peptides from Cor a 11 are the most suitable for the development of an LC-MS/MS method for the detection of hazelnut in food in terms of similarity in peptide intensity between different matrices and different hazelnut origins. However, peptides from Cor a 9 generally show a higher

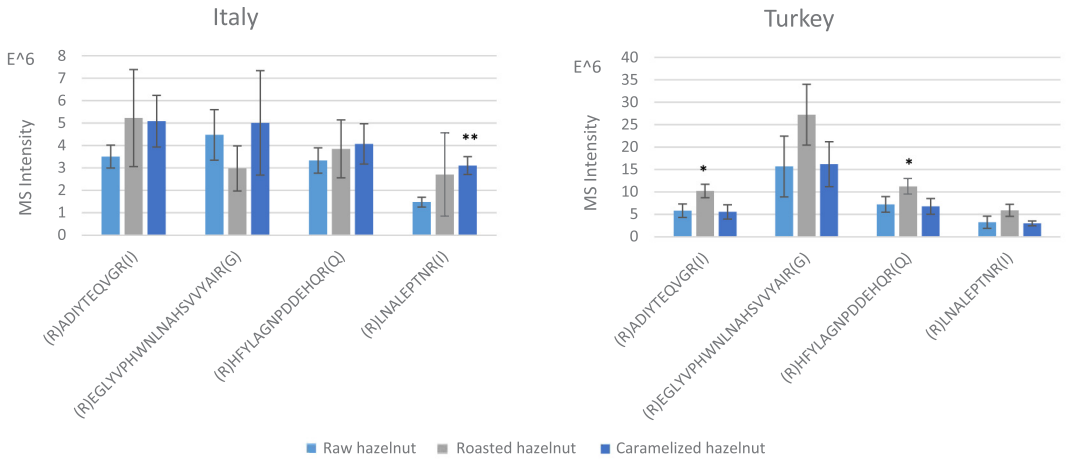



**Fig. 5.** Comparison of absolute MS intensities of tryptic peptides that were retained after several selection criteria from Cor a 9 (Q8W1C2) in three different test materials (non-processed hazelnut (Raw hazelnut), heated hazelnuts (Roasted hazelnut) and hazelnuts that have undergone induction of Maillard reactions (Caramelized hazelnut). MS signal intensities are expressed as the means of the peptide MS signal intensity ± 1 SD for 3 biological replicates. Significance levels were calculated by Student's *t*-test, comparing values from raw hazelnut to processed hazelnut (* = $p < 0.05$; ** = $p < 0.01$, ** = $p < 0.001$).

**Table 1**
Final selection of potential good biomarkers for the detection of hazelnut in food. Bold written peptides have not been published yet.

| Protein | Peptide | Position | Publications |
|---|---|---|---|
| Cor a 8 (Q9ATH2) | **AVNDASR** | 57–63 | |
| Cor a 9 (Q8W1C2) | LNALEPTNR | 43–51 | Planque et al. (2017); Sealey-Voyksner et al. (2016) |
| | **HFYLAGNPDDEHQR** | 195–208 | |
| | ADIYTEQVGR | 341–350 | Bignardi, Elviri, Penna, Careri, and Mangia (2010); Bignardi et al. (2013); Gu et al. (2018); Heick et al. (2011); Pilolli et al. (2017, 2018); Planque et al. (2019, 2017) |
| | **EGLYVPHWNLNAHSVVYAIR** | 377–396 | |
| Cor a 11 (Q8S4P9) | LLSGIENFR | 92–100 | Ansari et al. (2012); Costa et al. (2014) |
| | **ILQPVSAPGHFEAFYGAGGE DPESFYR** | 175–201 | |
| | **ALSQHEEGPPR** | 241–251 | |

absolute intensity, and thus, this protein is expected to be detectable in lower concentration of hazelnut in food compared to Cor a 11. Indeed, when developing highly sensitive tandem mass spectrometry detection methods, clearly those peptides giving the highest signals are prefered, this to reach the lowest limit of detection, which will be determined during validation of the method. Secondly, an LC-MS/MS method preferably contains multiple peptides per protein, and multiple proteins of the allergenic ingredient to strengthen reliability of the method. This implicates that the peptide from Cor a 8 on its own would not be ideal as biomarker being the only suitable peptide from this protein, and showing a low absolute intensity. Combining this peptide with peptides from Cor a 9 and Cor a 11 for developing an LC-MS/MS method will higher the confidence of the method. Because of the lack of research into hazelnut protein isoforms, and the small database of known hazelnut proteins, the list of selected peptides should be updated once more information is accessible in literature. Finally, the need of checking the performance of these peptides in different food matrices remains important.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work was supported by the Belgian Federal Public Service of Health, Food Chain Safety and Environment [RT 15/10 ALLERSENS].

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2019.125679.

## References

Ansari, P., Stoppacher, N., & Baumgartner, S. (2012). Marker peptide selection for the determination of hazelnut by LC-MS/MS and occurrence in other nuts. *Analytical and Bioanalytical Chemistry, 402*(8), 2607–2615. https://doi.org/10.1007/s00216-011-5218-6.

Bignardi, C., Elviri, L., Penna, A., Careri, M., & Mangia, A. (2010). Particle-packed column versus silica-based monolithic column for liquid chromatography–electrospray-linear ion trap-tandem mass spectrometry multiallergen trace analysis in foods. *Journal of Chromatography A, 1217*(48), 7579–7585. https://doi.org/10.1016/j.chroma.2010.10.037.

Bignardi, C., Mattarozzi, M., Penna, A., Sidoli, S., Elviri, L., Careri, M., & Mangia, A. (2013). A rapid size-exclusion solid-phase extraction step for enhanced sensitivity in multi-allergen determination in dark chocolate and biscuits by liquid chromatography-tandem mass spectrometry. *Food Analytical Methods, 6*(4), 1144–1152. https://doi.org/10.1007/s12161-012-9521-4.

Bonvehí, J. S. (1995). A chemical study of the protein fractions of Tarragona hazelnuts (Corylus avellana). *Zeitschrift Fur Lebensmittel-Untersuchung Und -Forschung, 201*(4), 371–374.

Costa, J., Ansari, P., Mafra, I., Oliveira, M. B. P. P., & Baumgartner, S. (2014). Assessing hazelnut allergens by protein- and DNA-based approaches: LC-MS/MS, ELISA and real-time PCR. *Analytical and Bioanalytical Chemistry, 406*(11), 2581–2590. https://doi.org/10.1007/s00216-014-7679-x.

Costa, J., Mafra, I., Carrapatoso, I., & Oliveira, M. B. P. P. (2016). Hazelnut allergens: Molecular characterization, detection, and clinical relevance. *Critical Reviews in Food Science and Nutrition, 56*(15), 2579–2605. https://doi.org/10.1080/10408398.2013.826173.

Cucu, T., Meulenaer, B. D., & Devreese, B. (2012). MALDI-based identification of stable hazelnut protein derived tryptic marker peptides. *Food Additives & Contaminants: Part A, 29*(12), 1821–1831. https://doi.org/10.1080/19440049.2012.719639.

Desjardins, P., Hansen, J. B., & Allen, M. (2009). Microvolume protein concentration determination using the NanoDrop 2000c spectrophotometer. Journal of Visualized

Experiments: JoVE, (33). doi:10.3791/1610.
Garino, C., Locatelli, M., Coïsson, J. D., D'Andrea, M., Cereti, E., Travaglia, F., & Arlorio, M. (2013). Gene transcription analysis of hazelnut (Corylus avellana L.) allergens Cor a 1, Cor a 8 and Cor a 11: A comparative study. *International Journal of Food Science & Technology, 48*(6), 1208–1217. https://doi.org/10.1111/ijfs.12079.
Gavage, M., Van Vlierberghe, K., Van Poucke, C., De Loose, M., Gevaert, K., Dieu, M., & Gillard, N. (2019). Selection of egg peptide biomarkers in processed food products by high resolution mass spectrometry. *Journal of Chromatography A, 1584*, 115–125. https://doi.org/10.1016/j.chroma.2018.11.036.
Godugu, B., Neta, P., Simón-Manso, Y., & Stein, S. E. (2010). Effect of N-terminal glutamic acid and glutamine on fragmentation of peptide ions. *Journal of the American Society for Mass Spectrometry, 21*(7), 1169–1176. https://doi.org/10.1016/j.jasms.2010.03.027.
Gu, S., Chen, N., Zhou, Y., Zhao, C., Zhan, L., Qu, L., & Ding, Y. (2018). A rapid solid-phase extraction combined with liquid chromatography-tandem mass spectrometry for simultaneous screening of multiple allergens in chocolates. *Food Control, 84*, 89–96. https://doi.org/10.1016/j.foodcont.2017.07.033.
Heick, J., Fischer, M., Kerbach, S., Tamm, U., & Popping, B. (2011). Application of a liquid chromatography tandem mass spectrometry method for the simultaneous detection of seven allergenic foods in flour and bread and comparison of the method with commercially available ELISA test kits. *Journal of AOAC International, 94*(4), 1060–1068.
Hustoft, H. K., Malerod, H., Wilson, S. R., Reubsaet, L., Lundanes, E., & Greibrokk, T. (2012). A critical review of trypsin digestion for LC-MS based proteomics. *Integrative Proteomics*. https://doi.org/10.5772/29326.
Jia, L., & Sun, Y. (2017). Protein asparagine deamidation prediction based on structures with machine learning methods. *PLOS ONE, 12*(7), e0181347. https://doi.org/10.1371/journal.pone.0181347.
Jonckheere, V., Fijałkowska, D., & Van Damme, P. (2018). Omics assisted N-terminal proteoform and protein expression profiling on methionine aminopeptidase 1 (MetAP1) deletion. *Molecular & Cellular Proteomics, 17*(4), 694–708. https://doi.org/10.1074/mcp.RA117.000360.
Lauer, I., Foetisch, K., Kolarich, D., Ballmer-Weber, B. K., Conti, A., Altmann, F., & Scheurer, S. (2004). Hazelnut (Corylus avellana) vicilin Cor a 11: Molecular characterization of a glycoprotein and its allergenic activity. *Biochemical Journal, 383*(2), 327–334. https://doi.org/10.1042/BJ20041062.
Manea, I., Ailenei, E., & Deleanu, D. (2016). Overview of food allergy diagnosis. *Clujul Medical, 89*(1), 5–10. https://doi.org/10.15386/cjmed-513.
Mariotti, F., Tomé, D., & Mirand, P. P. (2008). Converting nitrogen into protein—beyond 6.25 and Jones' factors. *Critical Reviews in Food Science and Nutrition, 48*(2), 177–184. https://doi.org/10.1080/10408390701279749.
McWilliam, V., Koplin, J., Lodge, C., Tang, M., Dharmage, S., & Allen, K. (2015). The prevalence of tree nut allergy: A systematic review. *Current Allergy and Asthma Reports, 15*(9), 54. https://doi.org/10.1007/s11882-015-0555-8.
Monaci, L., & Visconti, A. (2010). Immunochemical and DNA-based methods in food allergen analysis and quality assurance perspectives. *Trends in Food Science & Technology, 21*(6), 272–283. https://doi.org/10.1016/j.tifs.2010.02.003.
Neta, P., Pu, Q.-L., Kilpatrick, L., Yang, X., & Stein, S. E. (2007). Dehydration versus deamination of N-terminal glutamine in collision-induced dissociation of protonated peptides. *Journal of the American Society for Mass Spectrometry, 18*(1), 27–36. https://doi.org/10.1016/j.jasms.2006.08.016.
Nitride, C., Mamone, G., Picariello, G., Mills, C., Nocerino, R., Berni Canani, R., & Ferranti, P. (2013). Proteomic and immunological characterization of a new food allergen from hazelnut (Corylus avellana). *Journal of Proteomics, 86*, 16–26. https://doi.org/10.1016/j.jprot.2013.05.001.
Özenç, N., & Özenç, D. B. (2015). Nut traits and nutritional composition of hazelnut (Corylus avellana L.) as influenced by zinc fertilization. *Journal of the Science of Food and Agriculture, 95*(9), 1956–1962. https://doi.org/10.1002/jsfa.6911.
Pfeifer, S., Bublin, M., Dubiela, P., Hummel, K., Wortmann, J., Hofer, G., & Hoffmann-Sommergruber, K. (2015). Cor a 14, the allergenic 2S albumin from hazelnut, is highly thermostable and resistant to gastrointestinal digestion. *Molecular Nutrition & Food Research, 59*(10), 2077–2086. https://doi.org/10.1002/mnfr.201500071.
Pilolli, R., De Angelis, E., & Monaci, L. (2017). Streamlining the analytical workflow for multiplex MS/MS allergen detection in processed foods. *Food Chemistry, 221*, 1747–1753. https://doi.org/10.1016/j.foodchem.2016.10.110.
Pilolli, R., De Angelis, E., & Monaci, L. (2018). In house validation of a high resolution mass spectrometry Orbitrap-based method for multiple allergen detection in a processed model food. *Analytical and Bioanalytical Chemistry, 410*(22), 5653–5662. https://doi.org/10.1007/s00216-018-0927-8.
Planque, M., Arnould, T., Delahaut, P., Renard, P., Dieu, M., & Gillard, N. (2019). Development of a strategy for the quantification of food allergens in several food products by mass spectrometry in a routine laboratory. *Food Chemistry, 274*, 35–45. https://doi.org/10.1016/j.foodchem.2018.08.095.
Planque, M., Arnould, T., Dieu, M., Delahaut, P., Renard, P., & Gillard, N. (2016). Advances in ultra-high performance liquid chromatography coupled to tandem mass spectrometry for sensitive detection of several food allergens in complex and processed foodstuffs. *Journal of Chromatography. A, 1464*, 115–123. https://doi.org/10.1016/j.chroma.2016.08.033.
Planque, M., Arnould, T., Dieu, M., Delahaut, P., Renard, P., & Gillard, N. (2017). Liquid chromatography coupled to tandem mass spectrometry for detecting ten allergens in complex and incurred foodstuffs. *Journal of Chromatography. A, 1530*, 138–151. https://doi.org/10.1016/j.chroma.2017.11.039.
Poms, R. E., Klein, C. L., & Anklam, E. (2004). Methods for allergen analysis in food: A review. *Food Additives and Contaminants, 21*(1), 1–31. https://doi.org/10.1080/02652030310001620423.
Popping, B., & Diaz-Amigo, C. (2017). European regulations for labeling requirements for food allergens and substances causing intolerances: History and future. *Journal of AOAC International*. https://doi.org/10.5740/jaoacint.17-0381.
Prado, M., Ortea, I., Vial, S., Rivas, J., Calo-Mata, P., & Barros-Velázquez, J. (2016). Advanced DNA- and protein-based methods for the detection and investigation of food allergens. *Critical Reviews in Food Science and Nutrition, 56*(15), 2511–2542. https://doi.org/10.1080/10408398.2013.873767.
Ree, R., Varland, S., & Arnesen, T. (2018). Spotlight on protein N-terminal acetylation. *Experimental & Molecular Medicine, 50*(7), https://doi.org/10.1038/s12276-018-0116-z.
Rowley, E. R., Fox, S. E., Bryant, D. W., Sullivan, C. M., Priest, H. D., Givan, S. A., & Mockler, T. C. (2012). Assembly and characterization of the European hazelnut 'jefferson' transcriptome. *Crop Science, 52*(6), 2679–2686. https://doi.org/10.2135/cropsci2012.02.0065.
Sathe, S. K., Teuber, S. S., & Roux, K. H. (2005). Effects of food processing on the stability of food allergens. *Biotechnology Advances, 23*(6), 423–429. https://doi.org/10.1016/j.biotechadv.2005.05.008.
Sealey-Voyksner, J., Zweigenbaum, J., & Voyksner, R. (2016). Discovery of highly conserved unique peanut and tree nut peptides by LC-MS/MS for multi-allergen detection. *Food Chemistry, 194*, 201–211. https://doi.org/10.1016/j.foodchem.2015.07.043.
Wei, Y., Sathe, S. K., Teuber, S. S., & Roux, K. H. (2003). A Sensitive Sandwich ELISA for the Detection of Trace Amounts of Cashew (Anacardium occidentale L.) Nut in Foods. Journal of Agricultural and Food Chemistry, 51(11), 3215–3221. doi:10.1021/jf025977x.
Wilkins, M. R., Lindskog, I., Gasteiger, E., Bairoch, A., Sanchez, J. C., Hochstrasser, D. F., & Appel, R. D. (1997). Detailed peptide characterization using PEPTIDEMASS–a World-Wide-Web-accessible tool. *Electrophoresis, 18*(3–4), 403–408. https://doi.org/10.1002/elps.1150180314.

4.4. "Using High Resolution Mass Spectrometry for selecting peptide markers for the detection of the food allergen milk in food products" (Van Vlierberghe *et al*, manuscript submitted to Food Analytical Methods)

The strategy to identify potential peptide biomarkers was finally applied to milk by Kaatje Van Vlierberghe at ILVO.

Milk allergy is one of the most reported food allergies in Europe, especially in children (Schoemaker *et al*, 2015). Due to their high nutritional value and diverse functional properties, milk and milk proteins are used in multiple applications in modern food industry such as in meat product as meat protein substitutes or as fining agent in winemaking (Jana, 2017).

The developed and optimized sample preparation protocol was applied to processed milk matrices: freeze-dried whole milk as unprocessed matrix, freeze-dried Ultra High Temperature (UHT) treated milk as heated matrix and matrix subjected to Maillard reactions, fermented milk as low pH matrix and chocolate incurred with freeze-dried milk as fat-rich and complex matrix. Resulting peptides were analysed by HPLC-HRMS.

The identified peptides were filtered using a series of criteria to ensure method specificity, sensitivity and robustness. Thus, peptides must be specific for the allergenic ingredient, be robust to food processing, have originated from abundant proteins, be size-compatible with triple quadrupole mass spectrometry analysis, and not be prone to amino acid modifications or missed cleavages by the protease(s) used.

Considering these criteria, a list of 8 robust potential peptide biomarkers was obtained, allowing for the detection of the both whey and casein milk fractions. These peptide biomarkers were considered for the development of the quantitative UHPLC-MS/MS analysis method.

At the time of writing this thesis, the manuscript was submitted to Food Analytical Methods Journal.

# Using High Resolution Mass Spectrometry for selecting peptide markers for the detection of the food allergen milk in food products

Kaatje VAN VLIERBERGHE[1], Maxime GAVAGE[2,3], Marc DIEU[3], Patsy RENARD[3], Thierry ARNOULD[3], Nathalie GILLARD[2], Katleen COUDIJZER[1], Marc DE LOOSE[1], Kris GEVAERT[4,5], Christof VAN POUCKE*[1]

[1]ILVO Flanders research institute for agriculture, fisheries and food, Technology and Food Science Unit, Brusselsesteenweg 370, BE-9090 Melle, Belgium; kaatje.vanvlierberghe@ilvo.vlaanderen.be, christof.vanpoucke@ilvo.vlaanderen.be, marc.deloose@ilvo.vlaanderen.be, katleen.coudijzer@ilvo.vlaanderen.be

[2]CER Groupe, Rue du Point du Jour 8, 6900, Marloie, Belgium; m.gavage@cergroupe.be, n.gillard@cergroupe.be

[3]Laboratory of Biochemistry and Cell Biology (URBC)-Namur Research Institute for Life Sciences (NARILIS), University of Namur, 61, rue de Bruxelles, 5000 Namur, Belgium; marc.dieu@unamur.be, thierry.arnould@unamur.be, patsy.renard@unamur.be

[4]VIB-UGent Center for Medical Biotechnology, Albert Baertsoenkaai 3, BE-9000 Ghent, Belgium; kris.gevaert@vib-ugent.be

[5]Department of Biomolecular Medicine, Ghent University, Albert Baertsoenkaai 3, BE-9000 Ghent, Belgium

*** Corresponding author; telephone: +32 (0) 92723064; e-mail:** christof.vanpoucke@ilvo.vlaanderen.be; Brusselsesteenweg 370, BE-9090 Melle, Belgium

# 1. Abstract

Cow's milk allergy (CMA) is one of the most reported food allergies in Europe. To help patients suffering from food allergies, European legislation dictates the obligatory labeling of (derivatives of) 14 ingredients, among which is milk. Detection of milk in food is predominantly done by Enzyme Linked Immunosorbent Assay (ELISA) tests. Although sufficiently sensitive, such tests can be subjected to effects of food processing. An analytical method that is gaining interest in the field of allergen detection is Ultra-High Performance Liquid Chromatography (UHPLC) coupled to tandem Mass Spectrometry (MS), where the analyte is a target peptide. When such peptide biomarkers are selected with care, a food processing robust detection method can be developed. In this study, we produced milk-incurred food materials that underwent several processing techniques. For this, we carefully selected raw start material, worked in a fully controlled production setting and had exact knowledge of the food processing circumstances to ensure the trueness of processing applied to the material. This was followed by establishing tryptic peptide profiles from each matrix using UHPLC coupled to high resolution mass spectrometry (UHPLC-HRMS). Finally, a careful comparison of peptide profiles and intensities resulted in the selection of eight peptide biomarkers suitable for application in LC-MS/MS based milk detection methods, among which a α-lactalbumin specific peptide, which, for the first time, was determined to be stable in different incurred materials.

## Keywords

Milk, peptide biomarkers, selection, incurred test material, high resolution mass spectrometry, food allergen

# 2. Introduction

Food allergies are an important and a growing health problem worldwide, especially in industrialized countries (Costa et al. 2016). These immune-mediated responses to certain components (mostly proteins) of certain foods affect 17 million Europeans, which represents a doubling compared to 10 years ago (Manea et al. 2016). Cow's milk allergy (CMA) is one of the most reported food allergies in Europe (Schoemaker et al. 2015). In CMA, an immune reaction starts after consumption of milk or milk proteins in an IgE-mediated or a non-IgE-mediated manner, or a combination of both (Flom and Sicherer 2019). IgE-mediated reactions have an acute and rapid onset, while non-IgE mediated reactions are mostly long-term responses (Lifschitz and Szajewska 2015). The prevalence of CMA is age-dependent. Indeed, while 2-6% of young children are affected by CMA, only 0.1-0.5% of the adult population suffers from this condition (Flom and Sicherer 2019). Typical symptoms of IgE-mediated reactions to milk can be mild, such as vomiting and hives, but life-threatening responses such as an anaphylactic shock are also possible (Prasad et al. 2018). It is thus important to protect patients suffering from CMA. Since no cure for food allergy is currently available, complete avoidance of consuming milk or milk proteins is essential (Sicherer and Sampson 2014).

To help patients suffering from food allergies, European legislation dictates the obligatory labeling of (derivatives of) 14 ingredients (Popping and Diaz-Amigo 2017). Milk is one of these regulated culprit foodstuffs when used as an ingredient through regulation 1169/2011/CE. In modern food industry, milk and milk proteins are used in multiple applications because of their high nutritional value and diverse functional properties (Jana 2017). In fact, nothing is as widely used as a source of ingredients as dairy products (Fox 2001). Milk and milk proteins are for example often used in baked foodstuff and desserts, but also in meat product as meat protein substitutes. Indeed, given the higher nutritional value and lower energy content when compared to fat, dairy products made it to be ideal substituents in modern times in western countries where meat consumption tends to be reduced (Królczyk et al. 2016). Also, their ability to bind phenolic compounds makes them excellent fining agents for winemaking (Tolin et al. 2012). The use of milk derivatives and proteins in foods is also enhanced by the low cost for production companies (Liu et al. 2000). Hence, complete avoidance of milk is difficult in modern societies, which again highlights the need for accurate and correct labeling. In fact, not only patients suffering from CMA would benefit from correct labeling but patients with other pathologies that seem to have a link with milk intake such as diarrhea in patients suffering from irritable bowel syndrome, could benefit from milk (protein) avoidance (Cozma-Petruţ et al. 2017).

When deliberately added to food, labeling foods for the presence of milk (proteins) is obligated however, when the presence of milk (proteins) is accidental, clearly, this cannot be indicated. Through cross-contamination of shared equipment, during material handling and others, minute amounts of

milk can enter the final product (Taylor and Baumert 2010). As even trace amounts of milk (in the range of mg of milk proteins per kg of food matrix) can elicit a reaction in allergic individuals, such an accidental presence of milk (proteins) forms a health risk. Thus, to completely protect the allergic population, the presence of trace amounts of milk (proteins) should also be indicated. To achieve this, reliable methods for detecting milk or milk proteins in food are essential. Such methods must detect trace amounts of milk material to cover both deliberate and accidental presence of milk in food (Prado et al. 2016). Voluntary Incidental Trace Allergen Labelling (VITAL) has established a reference dose, defined as milligram protein level (total protein from an allergenic food) below which only the most sensitive individuals (1%) in an allergic population will experience an adverse reaction to 0.2 mg of total protein for milk (Westerhout et al. 2019), a value which should be reached by these detection methods. Detection of milk in food is predominantly done by Enzyme Linked Immunosorbent Assay (ELISA) tests, which are based on the recognition of an epitope present on a milk protein by an epitope-specific antibody. Although and sufficiently sensitive (up to < 0.3 ppm (Monaci et al. 2011a)), such tests can be subjected to effects of food processing. Indeed, proteins can undergo various modifications during food processing, such as denaturation and aggregation, and post-translational modifications (Taylor et al. 2009; Taylor and Baumert 2010; Prado et al. 2016). If the epitope recognized in an ELISA test is lost or altered during such food processing, underestimations of quantification or false negative results will be unavoidable. Another disadvantage of routine ELISA for detecting food allergens in general is the lack of a multiplexing ability (Cho et al. 2015). In terms of cost- and time-efficiency, methods that can detect multiple allergens during one analysis are thus desirable. In addition, as ELISA is based on epitope recognition, a risk of cross-reaction with an unknown substance remains, opening up the possibility for false positive results (Wei et al. 2003).

An analytical method that is gaining interest in the field of allergen detection is Ultra High Performance Liquid Chromatography (UHPLC) coupled to tandem Mass Spectrometry (MS/MS) (Pilolli et al. 2017). Here, the analyte is a protein- and ingredient-specific peptide, often derived through tryptic digestion of extracted proteins from the foodstuff. This method has several advantages, such as the possibility of multiplexing and higher accuracy. It also produces qualitative as well as quantitative results and gives exact information on the allergen. Another major advantage is its ability to be robust against modifications introduced by food processing when the peptide is selected with care. This particular advantage is essential for the purpose of a standardized test for a wide range of foodstuffs, since most food products in modern western countries are processed (Sathe and Sharma 2009). The peptides that are suitable to serve as biomarkers in a food processing-resistant and universally applicable detection method should meet several criteria. Such peptides are by high preference present in the digested protein extract in just one chemical form, thus containing no amino acids sensitive to processing-

induced, thus artificial post-translational modifications. Although some of these modifications can, at least to some extent, be predicted or anticipated (for example lysine glycation during heating of milk, and oxidation of methionine residues), it remains most efficient to experimentally determine those peptides that remain robust towards food processing.

This study aimed to identify and select suitable peptide markers robust toward food processing for the UHPLC-MS/MS based detection of milk allergens in food. To reach this objective, we produced in-house food materials incurred with milk that underwent several processing techniques. For this, we carefully selected raw start material, worked in a fully controlled production setting and had exact knowledge of the food processing circumstances to ensure the trueness of processing applied to the material. This step was followed by establishing tryptic profiles from each matrix using UHPLC-HRMS. Finally, a careful comparison of peptide profiles and intensities resulted in the selection of eight peptide biomarkers suitable for application in LC-MS/MS based milk detection methods.

# 3. Methods and Materials

## 3.1 Materials

All reagents and chemicals were of analytical grade unless stated otherwise.

Urea, tris(hydroxymethyl)aminomethane (Tris), tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl), iodoacetamide (IAA), tetraethylammonium bicarbonate (TEAB) and NAP-10 columns (GE Healthcare) were purchased from Sigma-Aldrich (Bornem, Belgium). Trifluoroacetic acid (TFA) was obtained from Thermo Fisher Scientific (Waltham, Massachusetts, USA), and solid-phase extraction cartridge (SEP) SEP-Pak C18 1 cc Vac Cartridge from Waters (Milford, Massachusetts, USA). Enzymatic digestion was performed with Trypsin Gold, Mass Spectrometry Grade from Promega (Madison, Wisconsin, USA), and Lysyl Endopeptidase, Mass Spectrometry Grade (Lys-C) from Wako Pure Chemical Industries, Ltd (Osaka, Japan). Acetonitrile, absolute methanol, 2-propanol (ULC/MS grade for all solvents), and formic acid were from Biosolve (Valkenswaard, the Netherlands). Ammonium phosphatide was kindly provided by Palsgaard A/S (Juelsminde – Denmark).

## 3.2 Production of test materials

Milk or processed milk incurred matrices were produced in the Food Pilot unit of ILVO. A batch of 100 L of raw whole cow's milk from ILVO dairy farm was used as start material. For storage and shelf-life purposes, 5 L of this raw whole milk was freeze-dried upon arrival, vacuum packed and stored for further use, and served as the unprocessed milk matrix. Total nitrogen content of this raw whole milk was determined by the Kjeldahl assay (Kjeldahl 1883). These values were multiplied with a value of 6.38 (Association of Official Analytical Chemists (AOAC) factor for milk) (Mariotti et al. 2008) to

estimate the protein content of this freeze-dried raw whole milk, resulting in a protein content of 24.73% .

***Preparation of a milk matrix subjected to heat treatment and Maillard reaction (Ultra High Temperature (UHT) milk):***

To induce the Maillard reaction (only occurring in the presence of reducing sugar, amino acids and heat), two types of Ultra High Temperature (UHT) treatments were applied on raw milk, namely direct steam infusion and indirect tubular heat exchange (UHT installation SPP, APV-SPX FLOW, Erpe-Mere, Belgium). For storage and shelf-life purposes, 5 L of each type of UHT milk was immediately freeze-dried and vacuum-packed upon further use.

***Preparation of fermented milk as a low pH milk matrix*:**

This matrix was produced by fermenting milk using a "set yoghurt" procedure. Raw whole milk was homogenized using a high-pressure homogenizer (Geo Niro, Hertogenbosch, The Netherlands) and heated to 45°C. Homogenized milk was then inoculated with yoghurt cultures (YO-MIX™ 860, St. thermophiles and Lb. bulgaricus, Danisco,) in a dosage of 200 CFU/1000 L. The inoculated milk was poured into plastic jars (100 g, DECA TP 69-155, DECA TP, Belgium), which were subsequently closed and incubated at 43°C. The pH was checked periodically until a pH of 4 was reached. The resulting yoghurt was stored at 4°C upon analysis, which was done within 10 days after production.

***Preparation of chocolate incurred with milk as a fatty rich milk matrix:***

Freeze-dried raw milk was milled and sieved into a fine powder (RETSCH sieve 500 µm). Secondly, chocolate was prepared analogous to chocolate production in previous work (Gavage et al. 2019, 2020; Van Vlierberghe et al. 2019). In brief, cacao butter and refiner flakes were melted and mixed with each other, after which 1.2% ammonium phosphatide was added. This mixture was stirred and, finally, to produce the milk-incurred chocolate, freeze-dried raw whole milk powder was mixed into the molten chocolate, resulting in a 38,500 ppm (mg milk proteins/kg chocolate) matrix. Liquid incurred chocolate was poured into chocolate molds, covered with plastic foil and left to cool down. Chocolate was kept overnight at 4 °C to cool further, set and solidify. The next day, chocolate bars were rasped into chocolate flakes. Chocolate flakes were kept in aliquots of 50 g and stored at 4°C in the dark. A proper blank chocolate was prepared following the same procedure, without adding the milk powder.

## 3.3 Sample preparation

Sample preparation is identical to the protocol described in our previous study (Van Vlierberghe et al. 2019). In short, proteins were extracted from food matrices, purified and digested with trypsin, after which peptides were purified. Each matrix was analyzed in biological triplicate.

***Protein extraction***

An amount of 1 g of matrix material was extracted with 10 mL extraction buffer (2 M urea, 200 mM Tris-HCl, pH 9.2) by shaking at 20 °C for 30 min at 300 rpm (Shaker plate SM 30 B, EDMUND BÜHLERBodelshausen, Germany) followed by 15 min sonication on ice at 100 % ultrasound power. The samples were centrifuged at 4,600 g for 10 min at 4 °C after which the middle liquid phase was transferred to a new tube. The protein concentration of the extract was subsequently determined with an A280 ultra-violet absorbance protein assay (NanoPhotometer® N60, Westburg, Leusden, The Netherlands) (Desjardins et al. 2009).

***Protein reduction, alkylation and purification***

Samples were diluted with extraction buffer to a protein concentration of 4 mg of proteins per mL. Next, proteins were reduced and alkylated (for 15 min at 37 °C in the dark) with TCEP-HCl (15 mM final concentration) and IAA (30 mM final concentration) under agitation at 500 rpm. The samples were desalted and buffer-exchanged by gel filtration chromatography (NAP-10 columns, GE Healthcare, United Kingdom). Proteins were eluted in 1.5 mL 6 M urea in 50 mM TEAB. The protein concentration was measured with the A280 ultra-violet absorbance protein assay.

***Enzymatic digestion***

An amount corresponding to 100 µg of proteins was transferred to a new micro centrifuge tube and diluted with 6 M urea in 50 mM TEAB to a final volume of 50 µL before addition of 2 µg of Lys-C, and digestion was performed for 2 h at 37 °C under 300 rpm agitation. The samples were diluted with 50 mM TEAB to a final volume of 300 µL, and 1 µg of trypsin was then added. Tryptic digestion was subsequently performed for 16 h at 37 °C under 300 rpm agitation. Digestion was stopped by adding TFA until a pH ≤ 3 was reached.

***Peptide purification***

Peptides were purified with C18 SPE cartridges. The cartridges were first washed with 1 mL acetonitrile/water (50/50, v/v) and equilibrated with 1 mL wash solvent (acetonitrile/water/TFA, 2/97.9/0.1, v/v/v). Samples were diluted 10 times with water, loaded onto the cartridges and washed with 1 mL of the same wash solvent. Peptides were eluted with 2 mL acetonitrile/water (80/20, v/v) and subsequently evaporated under a nitrogen flow at 40 °C. Peptide pellets were dissolved in 200 µL 0.1% formic acid and filtered (centrifugal filter unit 0.22 µm, Merck-Millipore, Carrigtwohill, Ireland).

## 3.4 Mass spectrometry

Solubilized peptides from a digest of 100 µg proteins were analyzed by high performance liquid chromatography coupled to high resolution mass spectrometry using a Waters Acquity UPLC® (Waters, Milford, MA VS) coupled to a SYNAPT G2-S High Definition Mass Spectrometer (TOF) (Waters). Tryptic peptides were chromatographically separated on an ACQUITY UPLC Peptide BEH C18 column (300 Å,

1.7 µm, 2.1 mm X 150 mm) (Waters) protected by an ACQUITY UPLC Peptide BEH C18 VanGuard Precolumn (300 Å, 1.7 µm, 2.1 mm X 5 mm) (Waters). Mobile phase A was 0.1% formic acid (solvent A) and mobile phase B was 0.1% formic acid in acetonitrile (solvent B). The solvent gradient was set as follows: initial: 99% solvent A, 1% solvent B; 60 min: 60% solvent A, 40% solvent B; 61 min: 15% solvent A, 85% solvent B, 65 min: 99% solvent A, 1% solvent B. The injection volume was set at 5 µL, the flow rate at 200 µL/min. Peptides were ionized in positive electrospray (ESI+) with the following instrument settings: capillary voltage 0.80 kV, sample cone voltage: 80 V, source temperature 120 °C and desolvation temperature: 400 °C. Data acquisition was done in continuous full scan mode in $MS^E$ mode (data independent acquisition) and the m/z range was 50-2,000. Fragmentation was done with collision induced dissociation, with a low trap collision energy for precursor ions of 4 eV and high collision energy ramping from 20 to 40 eV for fragment ions. Recording was done in resolution mode (20k FWHM). Prior to analysis, the HRMS was calibrated with sodium formate. Leucine enkephalin (200 pg/µL, mass 556.2771 Da, Waters) was used as lock mass, continuously infused and acquired (scan time of 0.2 s) with an interval of 30 s, but not yet applied for mass correction.

### 3.5 Data analysis

Raw HRMS data were analyzed with ProteinLynx Global Server™ (PLGS, Waters). Raw data were first automatically mass corrected using the lock mass recordings. Identified peaks were searched against a database containing the bovine proteome stored in Uniprot.org (*Bos Taurus, 6909 reviewed* entries), added with the 192 reviewed human keratin entries from Uniprot.org, the cationic bovine trypsin sequence, the Lysyl endopeptidase sequence from *Lysobacter enzymogenes* and the possible protein contaminants from cacao and bacteria from the fermentation culture YO-MIX™ 860 (*St. thermophiles* and *Lb. bulgaricus*). Search parameters were as follows: trypsin was chosen as the primary digestion enzyme and carbamidomethylation of cysteine residues was set as a fixed modification. Oxidation of methionine residues, conversion of N-terminal glutamine to pyroglutamate, and deamidation of asparagine and glutamine were set as variable modifications and 3 missed cleavages were allowed.

## 4. Results and discussion

### 4.1 Material selection and production

In order to select peptide biomarkers robust towards food processing, test materials containing milk that underwent multiple processing techniques are needed as the bases to finally yield peptide profiles. These materials were produced in-house, to have full control over the raw material, processing steps, cleanness of the equipment, environment and handling. The fact that materials are incurred is essential for the selection of peptide biomarkers, as in spiked processed material, proteins-of-interest are only added after processing, and will thus not have been subjected to any processing effects. Further, it is known that the milk protein content and distribution can vary depending on the

lactation stage, cattle breed, feed, individual animal genetics and other factors (Ng-Kwai-Hang et al. 1987; Heck et al. 2009; Franzoi et al. 2019). To have representative start material for producing all different milk matrices, we used a raw milk mixture from one milking day of several Holstein-Friesian breed animals of the ILVO dairy farm. By using this milk, we have full knowledge of feed, handling of the milk during milking and transport, cleanness of transportation vessels and others. We selected the Holstein-Friesian breed as it is the most used breed in Europe, outnumbering all other milk cattle breeds ([CSL STYLE ERROR: reference with no printed form.]). The preference for this breed is due to its large dairy production and has, in its liquid form, a typical protein content of 3.1% (Products 1988).

To analyze the effect of processing on the peptide profiles, different processed milk-containing test materials were produced using modern equipment on a semi-industrial scale. This resulted in the representative food matrices raw freeze-dried whole milk, freeze-dried UHT treated whole milk with the direct infusion technique, freeze-dried UHT treated whole milk with the indirect tubular exchange technique, yoghurt made from raw whole milk and chocolate incurred with freeze-dried raw whole milk.

## 4.2 Protein selection

Although total milk protein content is influenced by multiple factors, the protein percentage is generally considered to be 3.1% (Products 1988) This protein fraction can be divided into the casein proteins, accounting for 80% of the protein content and milk serum or whey proteins, accounting for about 20% of protein content (Davoodi et al. 2016). In this study, we focused on the most abundant milk proteins to select peptide biomarkers as these proteins are likely to yield the most intense peptide signals. Indeed, highly abundant peptides should contribute to higher sensitivities of detection when developing an LC-MS/MS method for detecting milk proteins in food. Although the full UniProtKB database for milk from *Bos taurus* (bovine) contains 99 entries, 95% of the true milk protein fraction is represented by just 6 proteins, of which 4 are casein proteins (α-S1-casein, α-S2-casein, κ-casein and β-casein), and two are major whey proteins (α-lactalbumin and β-lactoglobulin) (Wal 2002). After analysis of the HRMS data from digested protein extracts of freeze-dried raw whole milk powder, the 6 proteins resulting in the most intense peptides were the same as the 6 most abundant proteins in milk, presenting a first quality check on the data. These proteins were also found in all the processed test materials.

## 4.3 Peptide biomarker selection

When detecting food allergens using LC-MS/MS, the target analytes are peptides derived from extracted proteins through digestion by a protease, mostly trypsin. Trypsin hydrolyzes peptide bonds C-terminal to arginine and lysine residues (except when these residues are followed by a proline or when lysine has undergone some post-translational modifications) with high specificity (Hustoft et al.

2012), and therefore renders many peptides of an ideal length (thus, on average 10-20 amino acids), basicity and hydrophobicity for LC-MS/MS. The lysine and arginine side-chains indeed makes most tryptic peptides rather basic, thus perfect for ionization in the ion source of the mass spectrometer. For each of the milk matrices, tryptic peptide profiles were created by analysis of HRMS data. Peptides were considered identified when they were detected in all three biological repeats with high confidence (score 2 in ProteinLynx, as analysed with ProteinLynx GlobalSERVER V3.0.3 (Waters Corporation)). Parallel to this, an *in silico* tryptic digestion using PeptideMass ([CSL STYLE ERROR: reference with no printed form.]) rendered theoretical tryptic peptide profiles for the 6 proteins selected. Here, no missed cleavages were allowed because in further downstream selection, only peptides without missed cleavages will be withheld, this to eliminate the chance of underestimation of the true peptide intensity because of possible multiple chemical forms of the peptide. Peptides minimally consisting of 7 amino acids were considered for specificity reasons. This criterium is also applied by the Allergen Peptide Browser, a database for LC-MS/MS based food allergen detection publications and peptide biomarkers used in those studies (Croote and Quake 2016). HRMS results from milk extracts were compared with this theoretical, *in silico* generated list of peptides. In freeze-dried raw whole milk, we were able to detect 8 out of 10 possible peptides for α-S1-casein, 10 out of 13 for α-S2-casein, 5 out of 7 for β-casein, 4 out of 5 for κ-casein, 4 of 5 for α-lactalbumin and 9 out of 10 possible peptides for β-lactoglobulin **(Table 1)**.

#### 4.3.1 Peptide specificity to bovine milk

Suitable peptide biomarkers for detection of milk in food should be specific for milk from the mammary gland of farmed animals as stated in the European legislation. The general description of milk from multiple farmed animals arises from the fact that milk proteins from different animals show large homology, and that patients suffering from cow's milk allergy are often also susceptible towards other animal milks (Restani et al. 1999). To check specificity, each peptide was analyzed by BLAST (Basic Local Alignment Search Tool) by searching the sequence against the full UniProtKB database. To be specific, in previous works, we used the criterion that a peptide could only fully match to the sequence of the corresponding proteins from the allergen itself (Gavage et al. 2019, 2020; Van Vlierberghe et al. 2019). However, the large homology was also reflected in this search results, as all detected milk peptides showed full identify to proteins from animal species as well. Based on the large amount of matched peptides to other organisms, we used the criteria for specificity applied on the allergen peptide browser site, thus not matching to any other allergens included in European legislation (Croote and Quake 2016), and to no other food ingredients . From the 6 proteins included in our selection, no peptide fully matched to other food allergens and thus were considered milk-specific.

### 4.3.2 Peptide sequence-based selection

*Common amino acid modifications*

In order to maximize the accuracy when using the selected peptide biomarkers, peptides containing certain amino acids were excluded, being methionine, cysteine and N-terminal glutamine. Indeed, for quantification purposes, the signal intensities of the biomarker peptides should reflect the abundance of milk proteins in the food product. Such a statement can only be fulfilled when only one chemical form of the peptide is present in the digested protein extract of the food sample. Multiple forms of a peptide can occur when amino acids prone to modifications are present in the peptide sequence. Such modifications are for example methionine oxidation and cysteine carbamidomethylation, resulting in a mass shift of +15.995 Da and +57.0520 Da respectively. Methionine oxidation is an artificial modification that readily occurs in an uncontrollable manner when proteins are exposed to air. Although one reduces and alkylates cysteine residues during sample preparation, depending on the food matrix, this reaction can be suboptimal, resulting in a risk of erroneous quantification of the peptide when using it as a biomarker, resulting in erroneous quantification of the allergen itself. In addition, the N-terminal glutamine to pyroglutamic acid conversion (mass shift of -17.027 Da) also occurs, again resulting in multiple forms of a given peptide (Neta et al. 2007). Moreover, pyroglutamic acid is known to negatively affect peptide fragmentation, and sub-efficient fragmentation could reduce detectability of the peptide in LC-MS/MS detection methods (Godugu et al. 2010) (**Table 1**).

*Motifs and sequences*

Based on the intrinsic properties of trypsin, peptides containing KP or RP motifs are avoided. Although trypsin is known not to cleave the peptide after K or R if followed by a P residue, this occasionally does happen (Rodriguez et al. 2008). Peptides carrying other types of missed cleavages were also excluded. For example, the motifs KK, KR, RR or RK at the beginning or end of a peptide sequence often result in two detected types of the peptide as peptides can be cleaved both after the first as well as after the second K or R. Further, it is known that the immediate proximity of a negatively charged residue (e.g. D or E) can result in partial proteolysis by trypsin (Boyle et al. 1991). Indeed, this is what we observed in 6 peptides **(Figure 1).**

One should also consider that motifs such K/R – X – pSer/pThr can also result in suboptimal tryptic digestion and, this especially for caseins, which are known to hold multiple phosphoserines. Such missed cleavages however, were not detected in our data.

The peptide HPHPHLSFMAIPPK from κ-casein was excluded because of the $F_{119}M_{132}$ motif known to be cleaved *in vivo* through chymosin activity (Reid et al. 1997). This occurs for example during cheese making, forming again multiple variants of the peptide and thus potentially diminishing the intensity of the tryptic peptide.

Other motifs and sequences that were avoided are the NG-motif, because of the tendency of N becoming deaminated (Jia and Sun 2017), and sequences at the N-terminus of the protein because of the risk of holding a signal peptide which is cleaved from the proteins destined for the secretory pathway. Moreover, protein N-termini can undergo excision of the N-terminal initiator methionine (iMet) residue in nascent polypeptide chains (Jonckheere et al. 2018) and N-terminal acetylation(Ree et al. 2018), influencing detection of these peptides **(Table 1)**.

*Phosphorylation or serine residues in caseins*

After synthesis of α-S1-casein, α-S2-casein, β-casein, κ-casein in the Golgi apparatus, different serine residues can be phosphorylated by protein kinases (Bingham et al. 1972). These phosphorylation events are required for the formation of casein micelles through interaction with calcium phosphate, important for the calcium binding property of milk and the health benefits associated to it (Li et al. 2012). Different phosphorylation proteoforms can exist for each casein protein: for κ-casein, 3 isoforms are known: κ-casein 1P, 2P and 3P, on residues $S_{148,170}$ and $_{187}$. β-caseins are commonly found in 2 phosphorylation isoforms, β-casein 4P, phosphorylated on $S_{30,\ 32,\ 33}$ and $_{34}$ and β-casein 5P, additionally phosphorylated on $S_{50}$, which is only found in certain genetic variants. Phospho-isoforms of α-S1-casein vary between 8P and 9P, on $S_{56,\ 61,\ 63,\ 79,\ 81,\ 82,\ 83,\ 90}$ and $_{130}$. α-S2-casein is present in phosphorylation isoforms from 10P to 13P $S_{23,24,25,28,46,71,72,73,76,144,145,150}$ and $_{158}$ (Fang et al. 2016). However, these degrees of casein phosphorylations can vary, even between individual cows (Poulsen et al. 2016). Since multiple phosphorylation proteoforms can be present, the true abundance of a given peptide cannot simply be derived from one peptide intensity when a phosphorylated serine is located in this peptide sequence. Moreover, the acidic phosphate groups are known to interfere with peptide ionization and fragmentation, resulting in suboptimal detection and thus an underestimation of the peptide (Paradela and Albar 2008), clearly undesirable for quantification purposes. To select peptide biomarkers that are fully robust and resistant to any form of variation, peptides containing serine residues that can be phosphorylated should therefore be excluded. For κ-casein, no peptides were excluded as the three serines that are potentially phosphorylated were not covered by the tryptic peptides (**Table 1).**

*Glycosylations of milk proteins*

Some milk proteins have known glycosylation sites, for example the glycosylation of $N_{64}$ and $N_{90}$ in α-lactalbumin (Hochwallner et al. 2010). However, these sites are not always glycosylated. Indeed, $N_{64}$ of α-lactalbumin is known to be glycosylated in only 10% of all α-lactalbumin present in milk (Slangen and Visser 1999). Also, κ-casein has multiple glycosylation sites in its C-terminal part, but only 40% of the κ-casein present in milk is known to be glycosylated (Holland et al. 2004). Again, the abundance of

a peptide cannot be correctly estimated if this peptide contains (a) possible glycosylation site(s). Therefore, peptides containing such sites were also excluded (**Table 1).**

After applying the above selection criteria, the following number of peptides were retained for further selection: 4 peptides from α-S1-casein, 4 peptides for α-S2-casein, 3 peptides for β-casein, 2 for κ-casein, 1 for α-lactalbumin and 2 for β-lactoglobulin **(Table 1, rows in bold)**.

### 4.3.4 Selection based on robustness during food processing

Up to this stage, peptides were selected based on their composition and thus intrinsic properties, and our experimental observations. Further selection of peptide biomarkers was based on their robustness towards food processing. Therefore, absolute peptide intensities were compared between the different milk-containing matrices. A peptide is considered stable and thus robust towards food processing when its intensity in the different matrices is not significantly different. However, this can only be evaluated if equal amounts of milk proteins are digested and equal amounts of milk peptides are analyzed by HRMS. Since milk yoghurt and milk-incurred chocolate also contain bacterial and chocolate proteins, the true amount of milk proteins digested and the true amount of peptides analyzed with HRMS remain elusive. We therefore first evaluated peptide detection robustness towards heating and the Maillard reaction, after which the influence of low pH and a fatty rich environment was also evaluated.

*Robustness towards heat treatment*

Heat treatment of milk will induce the Maillard reaction with proteins, a non-enzymatic glycation between an amino group and a carbonyl group, mostly from a reducing sugar present in the food (Choudhary et al. 2017). Reducing sugars are always present in milk in the form of lactose. The most susceptible groups to undergo glycation are lysine side-chains, followed by arginine side-chains (Münch et al. 1999). Through such glycation, occurring during heat treatment of milk, trypsin will no longer recognize these modified lysine and arginine residues as cleavage sites, resulting in a suboptimal digestion. As a heating process, two types of UHT treatment often applied to sterilize milk in western countries were chosen. These being direct UHT in which milk is heated for a short time by injection of a heat steam, and indirect UHT during which milk is heated for a longer time through indirect heat exchange, for example from parallel tubes. The heat load is higher in indirect UHT, with a heat transfer time of 10 s and more, compared to direct UHT, where the heat transfer time is less than 1 s, and protein modifications are expected to be more pronounced with higher heat loads (Datta et al. 2002). Next to the heat load dependency, glycations are also protein dependent, where whey proteins are much more susceptible for the Maillard reaction compared to caseins (Cardoso et al. 2018). Two peptides from β-lactoglobulin (a whey protein) and 3 peptides from caseins (β-casein and α-S1-casein) showed missed cleavages, and when the percentages of the missed cleaved versions of these peptides

were compared over freeze-dried whole milk (unprocessed milk), freeze-dried UHT milk in a direct (steam injection) and indirect (tubular heat exchange) way, the whey proteins showed a significant difference in missed cleaved portion when heat was applied over a longer period of time compared to the raw milk or a short application of heat, while in the case of caseins, there was no significant difference **(Figure 2)**. In the end, 8 of the 13 peptides were found to be robust against UHT treatment **(Figure 3)**.

*Robustness in a low pH or in a fat-rich environment*

Robustness towards incurring in chocolate (fat-rich environment) and fermentation (to yoghurt, low pH environment) could only be evaluated by visual comparison of absolute intensities. Indeed, the presence of chocolate or bacterial proteins would result in a lower estimated peptide intensity when equal amounts of extracted proteins are digested and analyzed. Thus, a lower peptide intensity does not directly correlate with sensitivity towards food processing. To give an overview of the influence of fat-rich environment and low pH, **Figure 4** shows the absolute peptide intensities of the UHT robust selected peptides in all four materials. In general, the α-casein peptides and the peptide from β-lactoglobulin show good robustness towards a fat-rich environment, while the κ-casein peptides and the peptide from α-lactalbumin were found to be more sensitive. Incurring the milk in a low pH environment (whole milk fermentation into yoghurt) seems to strongly affect the detectability of peptides in general, resulting in relatively very low intensities or even no detection. This could partially be explained by the fact that at low pH, milk proteins denature and aggregate, and thus become more difficult to extract and subsequently digest. However, it is also known that the bacteria used as starters for milk fermentation hydrolyze the milk proteins, resulting in lower amounts of proteins, and thus lower peptide intensities (Pescuma et al. 2007; Paul and Somkuti 2009).

# 5. Discussion

Based on our findings, 8 peptides robust towards the induction of the Maillard reaction and heat treatment were identified, 5 of which showed good stability in chocolate as well. To our knowledge, this is the first systematic and experimentally based approach for the selection of suitable milk peptide biomarkers robust towards multiple, often applied food processing techniques. The use of in-house produced, incurred test material combined with the full scan analysis by high resolution mass spectrometry are major advantages with regards to building peptide databases for each type of milk-containing food matrix.

The identified candidate peptides are, as expected, all previously reported as biomarker in other LC-MS/MS based milk detection methods. Indeed, a recent comprehensive review on proteotypic peptide markers for detection of six major food allergens showed that milk as an allergen is the most investigated source of food allergen, and more importantly, has the highest consensus in selected

signature peptides across different works ([CSL STYLE ERROR: reference with no printed form.]). Many of these markers have already been validated in several food materials, but often in spiked material, and no other paper reported on the use of such a comprehensive and standardized approach for a selection of robust peptides in multiple matrices simultaneously.

Strikingly, some peptides that are very frequently used as marker for milk detection are excluded from our selection. For example, the most cited peptide, FFVAVPFPEVFGK, cited in 25 publications (Monaci et al. 2010a, b, 2011b, 2013, 2014; Ansari et al. 2011; Heick et al. 2011b, a; Newsome and Scholl 2013; Losito et al. 2013; Pilolli et al. 2014, 2017, 2018; Parker et al. 2015; Lamberti et al. 2016; Planque et al. 2016, 2017b, a, 2019; Ke et al. 2017; De Angelis et al. 2017; Boo et al. 2018; Groves et al. 2018; Gu et al. 2018; Montowska and Fornal 2019; Qi et al. 2019), in fact comes with the risks of underestimating milk in food given the negatively charged glutamic acid residue close to its terminal lysine, potentially hindering cleavage at this site (Boyle et al. 1991). Other frequently cited peptides excluded based on sequence specific properties were DMPIQAFLLYQEPVLGPVR (Monaci et al. 2010a, 2013; Cereda et al. 2010; Tolin et al. 2012; Chen et al. 2015; Montowska and Fornal 2018) from β-casein and the peptides VLVLDTDYK (Figeys et al. 1996; Lutter et al. 2011; Parker et al. 2015; Planque et al. 2016, 2017a, 2019), LSFNPTQLEEQCHI (Parker et al. 2015; Ke et al. 2017; Planque et al. 2017a; Montowska and Fornal 2018) and VYVEELKPTPEGDLEILLQK (Figeys et al. 1996; Planque et al. 2016, 2017a; Montowska and Fornal 2018) from β-lactoglubulin, of which the reason for exclusion is given in Table 1. However, some frequently cited peptides were excluded based on their susceptibility towards food processing as well, being the peptides HQGLPQEVLNENLLR (Lee and Kim 2010; Monaci et al. 2010a, 2011b, 2013; Cereda et al. 2010; Tolin et al. 2012; Newsome and Scholl 2013; Mattarozzi et al. 2014; Gomaa and Boye 2015; Parker et al. 2015; Planque et al. 2016, 2017a; Montowska and Fornal 2018; Gu et al. 2018) from α-S1-casein and GPFPIIV (Monaci et al. 2010a, 2011b, 2013; Cereda et al. 2010; Ansari et al. 2011; Tolin et al. 2012; Losito et al. 2013; Chen et al. 2015), VLPVPQK (Monaci et al. 2010a; Ansari et al. 2011; Tolin et al. 2012; Losito et al. 2013; Mattarozzi et al. 2014; Chen et al. 2015; Gu et al. 2018) and AVPYPQR (Cereda et al. 2010; Lutter et al. 2011; Tolin et al. 2012; Losito et al. 2013; Mattarozzi et al. 2014; Chen et al. 2015; Gu et al. 2018) from β-casein. On the contrary, the peptide VGINYWLAHK from α-lactalbumin seems a good candidate, showing certain robustness towards matrices containing high amounts of fats and polyphenols. However, this peptide has only been used in two other studies (Ansari et al. 2011; Zhang et al. 2012), of which none analyzed its behavior in detail in both incurred thermally processed and complex fatty rich matrix. The work of Zhang *et al*. describes the use of a synthetic version of a marker peptide as an internal standard for LC-MS/MS based quantification, and the availability of such an internal standard could be an additional advantage for further development of an LC-MS/MS based detection method. The fact that the peptide in this work showed excellent

robustness towards denaturation of α-lactalbumin, combined with our findings of robustness towards the Maillard reaction and the presence of polyphenols and fats, strengthens the choice of this peptide as a possible biomarker for detection of milk. Moreover, as most milk detection methods use peptides from highly abundant caseins, products contaminated with only whey proteins might be identified as false negatives. In fact, not a single peptide from α-lactalbumin has been proven to be stable in thermally processed or chocolate-based incurred matrices, so we here propose the first experimentally determined peptide in this manner. As α-lactalbumin is known for its various physiologic and nutritional functionalities, and its well-balanced supply of essential amino acids, isolates of this protein are indeed frequently used in the food industry (Layman et al. 2018), inherently forming a risk for cross contamination. The importance of including peptide biomarkers from this protein in a milk detection method are thus evident.

A low pH environment (yoghurt samples) was found to significantly influence peptide intensities however, care should be taken given the uncertainty of bacterial proteins that were co-extracted and digested, which can reduce the intensity of the milk peptides. Further, fermenting bacteria may also hydrolyze and thus reduce the levels of milk proteins.

The 8 peptides showed overall good intensities, except for one peptide from α-s2-casein. However, they were not the most intense peptides identified. Multiple peptides with higher absolute intensities were excluded from the final selection based on their intrinsic properties, amino acid composition and/or experimental findings. In fact, when considering the peptides showing the highest intensity for each milk-containing matrix, only 3 of 8 selected peptides make it to the top 20 (**Figure 5**). Four peptides in particular were the most intense in all materials, being (R)DMPIQAFLLYQEPVLGPVR(G), (R)GPFPIIV(-), (R)FFVAPFPEVFGK(E) and (K)HQGLPQEVLNENLLR(F). Strikingly, the absolute intensities of the 3 peptides from our selection included in the top 20 represent only 20-30% of the absolute intensity of the four high abundant, but non-selected peptides, and this trend was pronounced in all matrices. We propose therefore to keep these 4 highly intense peptides as adequate candidates for screening indication of milk in (processed) food, and use the 8 stable peptides we selected for quantification.

# 6. Conclusion

In this work, we describe the systematic and experimentally based approach for the selection of suitable milk peptide biomarkers robust towards multiple, often applied food processing techniques, making use of in-house produced, incurred test material combined with full scan analysis by high resolution mass spectrometry. Eventually 8 peptides robust towards the induction of the Maillard reaction and heat treatment were identified, 5 of which showed good stability in chocolate as well.

# 7. Compliance with Ethical Standards


## Funding:

This study was funded by the Belgian Federal Public Service of Health, Food Chain Safety and Environment [ RT 15/10 ALLERSENS].


## Conflicts of interest

The authors declare that they have no conflict of interest.

## Ethical approval:

This article does not contain any studies with human participants or animals performed by any of the authors.

# 8. References


Ansari P, Stoppacher N, Rudolf J, et al (2011) Selection of possible marker peptides for the detection of major ruminant milk proteins in food by liquid chromatography-tandem mass spectrometry. Anal Bioanal Chem 399:1105–1115. https://doi.org/10.1007/s00216-010-4422-0

Bingham EW, Farrell HM, Basch JJ (1972) Phosphorylation of casein. Role of the golgi apparatus. J Biol Chem 247:8193–8194

Boo CC, Parker CH, Jackson LS (2018) A Targeted LC-MS/MS Method for the Simultaneous Detection and Quantitation of Egg, Milk, and Peanut Allergens in Sugar Cookies. J AOAC Int 101:108–117. https://doi.org/10.5740/jaoacint.17-0400

Boyle WJ, van der Geer P, Hunter T (1991) Phosphopeptide mapping and phosphoamino acid analysis by two-dimensional separation on thin-layer cellulose plates. Methods Enzymol 201:110–149. https://doi.org/10.1016/0076-6879(91)01013-r

Cardoso HB, Wierenga PA, Gruppen H, Schols HA (2018) Maillard induced glycation behaviour of individual milk proteins. Food Chem 252:311–317. https://doi.org/10.1016/j.foodchem.2018.01.106

Cereda A, Kravchuk AV, D'Amato A, et al (2010) Proteomics of wine additives: mining for the invisible via combinatorial peptide ligand libraries. J Proteomics 73:1732–1739. https://doi.org/10.1016/j.jprot.2010.05.010

Chen Q, Zhang J, Ke X, et al (2015) Quantification of bovine β-casein allergen in baked foodstuffs based on ultra-performance liquid chromatography with tandem mass spectrometry. Food Addit Contam Part Chem Anal Control Expo Risk Assess 32:25–34. https://doi.org/10.1080/19440049.2014.990994

Cho CY, Nowatzke W, Oliver K, Garber EAE (2015) Multiplex detection of food allergens and gluten. Anal Bioanal Chem 407:4195–4206. https://doi.org/10.1007/s00216-015-8645-y

Choudhary S, Arora S, Kumari A, et al (2017) Effect of quality of milk on maillard reaction and protein oxidation during preparation of cow and buffalo milk khoa. J Food Sci Technol 54:2737–2745. https://doi.org/10.1007/s13197-017-2710-9

570 Costa J, Mafra I, Carrapatoso I, Oliveira MBPP (2016) Hazelnut Allergens: Molecular Characterization,
571 Detection, and Clinical Relevance. Crit Rev Food Sci Nutr 56:2579–2605.
572 https://doi.org/10.1080/10408398.2013.826173

573 Cozma-Petruţ A, Loghin F, Miere D, Dumitraşcu DL (2017) Diet in irritable bowel syndrome: What to
574 recommend, not what to forbid to patients! World J Gastroenterol 23:3771–3783.
575 https://doi.org/10.3748/wjg.v23.i21.3771

576 Croote D, Quake SR (2016) Food allergen detection by mass spectrometry: the role of systems
577 biology. Npj Syst Biol Appl 2:. https://doi.org/10.1038/npjsba.2016.22

578 Datta N, Elliott AJ, Perkins ML, Deeth HC (2002) Ultra-high-temperature (UHT) treatment of milk:
579 Comparison of direct and indirect modes of heating. Aust J Dairy Technol Melb 57:211–227

580 Davoodi SH, Shahbazi R, Esmaeili S, et al (2016) Health-Related Aspects of Milk Proteins. Iran J Pharm
581 Res IJPR 15:573–591

582 De Angelis E, Pilolli R, Monaci L (2017) Coupling SPE on-line pre-enrichment with HPLC and MS/MS
583 for the sensitive detection of multiple allergens in wine. Food Control 73:814–820.
584 https://doi.org/10.1016/j.foodcont.2016.09.031

585 Desjardins P, Hansen JB, Allen M (2009) Microvolume protein concentration determination using the
586 NanoDrop 2000c spectrophotometer. J Vis Exp JoVE. https://doi.org/10.3791/1610

587 Fang ZH, Visker MHPW, Miranda G, et al (2016) The relationships among bovine αS-casein
588 phosphorylation isoforms suggest different phosphorylation pathways. J Dairy Sci 99:8168–
589 8177. https://doi.org/10.3168/jds.2016-11250

590 Figeys D, van Oostveen I, Ducret A, Aebersold R (1996) Protein identification by capillary zone
591 electrophoresis/microelectrospray ionization-tandem mass spectrometry at the
592 subfemtomole level. Anal Chem 68:1822–1828

593 Flom JD, Sicherer SH (2019) Epidemiology of Cow's Milk Allergy. Nutrients 11:.
594 https://doi.org/10.3390/nu11051051

595 Fox PF (2001) Milk proteins as food ingredients. Int J Dairy Technol 54:41–55.
596 https://doi.org/10.1046/j.1471-0307.2001.00014.x

597 Franzoi M, Niero G, Visentin G, et al (2019) Variation of Detailed Protein Composition of Cow Milk
598 Predicted from a Large Database of Mid-Infrared Spectra. Anim Open Access J MDPI 9:.
599 https://doi.org/10.3390/ani9040176

600 Gavage M, Van KV, Van CP, et al (2020) High-resolution mass spectrometry-based selection of peanut
601 peptide biomarkers considering food processing and market type variation. Food Chem
602 304:125428–125428. https://doi.org/10.1016/j.foodchem.2019.125428

603 Gavage M, Van Vlierberghe K, Van Poucke C, et al (2019) Selection of egg peptide biomarkers in
604 processed food products by high resolution mass spectrometry. J Chromatogr A 1584:115–
605 125. https://doi.org/10.1016/j.chroma.2018.11.036

606 Godugu B, Neta P, Simón-Manso Y, Stein SE (2010) Effect of N-Terminal Glutamic Acid and Glutamine
607 on Fragmentation of Peptide Ions. J Am Soc Mass Spectrom 21:1169–1176.
608 https://doi.org/10.1016/j.jasms.2010.03.027

609 Gomaa A, Boye J (2015) Simultaneous detection of multi-allergens in an incurred food matrix using
610 ELISA, multiplex flow cytometry and liquid chromatography mass spectrometry (LC–MS).
611 Food Chem 175:585–592. https://doi.org/10.1016/j.foodchem.2014.12.017

612 Groves K, Cryar A, Walker M, Quaglia M (2018) Assessment of Recovery of Milk Protein Allergens
613 from Processed Food for Mass Spectrometry Quantification. J AOAC Int 101:152–161.
614 https://doi.org/10.5740/jaoacint.17-0214

615 Gu S, Chen N, Zhou Y, et al (2018) A rapid solid-phase extraction combined with liquid
616 chromatography-tandem mass spectrometry for simultaneous screening of multiple
617 allergens in chocolates. Food Control 84:89–96.
618 https://doi.org/10.1016/j.foodcont.2017.07.033

619 Heck JML, Schennink A, van Valenberg HJF, et al (2009) Effects of milk protein variants on the protein
620 composition of bovine milk. J Dairy Sci 92:1192–1202. https://doi.org/10.3168/jds.2008-1208

621 Heick J, Fischer M, Kerbach S, et al (2011a) Application of a liquid chromatography tandem mass
622 spectrometry method for the simultaneous detection of seven allergenic foods in flour and
623 bread and comparison of the method with commercially available ELISA test kits. J AOAC Int
624 94:1060–1068

625 Heick J, Fischer M, Pöpping B (2011b) First screening method for the simultaneous detection of seven
626 allergens by liquid chromatography mass spectrometry. J Chromatogr A 1218:938–943.
627 https://doi.org/10.1016/j.chroma.2010.12.067

628 Hochwallner H, Schulmeister U, Swoboda I, et al (2010) Visualization of clustered IgE epitopes on α-
629 lactalbumin. J Allergy Clin Immunol 125:1279-1285.e9.
630 https://doi.org/10.1016/j.jaci.2010.03.007

631 Holland JW, Deeth HC, Alewood PF (2004) Proteomic analysis of kappa-casein micro-heterogeneity.
632 Proteomics 4:743–752. https://doi.org/10.1002/pmic.200300613

633 Hustoft HK, Malerod H, Wilson SR, et al (2012) A Critical Review of Trypsin Digestion for LC-MS Based
634 Proteomics. Integr Proteomics. https://doi.org/10.5772/29326

635 Jana A (2017) High Protein Milk Ingredients - A Tool for Value-Addition to Dairy and Food Products. J
636 Dairy Vet Anim Res 6:. https://doi.org/10.15406/jdvar.2017.06.00171

637 Jia L, Sun Y (2017) Protein asparagine deamidation prediction based on structures with machine
638 learning methods. PLOS ONE 12:e0181347. https://doi.org/10.1371/journal.pone.0181347

639 Jonckheere V, Fijałkowska D, Van Damme P (2018) Omics Assisted N-terminal Proteoform and
640 Protein Expression Profiling On Methionine Aminopeptidase 1 (MetAP1) Deletion. Mol Cell
641 Proteomics 17:694–708. https://doi.org/10.1074/mcp.RA117.000360

642 Ke X, Zhang J, Lai S, et al (2017) Quantitative analysis of cow whole milk and whey powder
643 adulteration percentage in goat and sheep milk products by isotopic dilution-ultra-high
644 performance liquid chromatography-tandem mass spectrometry. Anal Bioanal Chem
645 409:213–224. https://doi.org/10.1007/s00216-016-9987-9

646 Kjeldahl J (1883) Neue Methode zur Bestimmung des Stickstoffs in organischen Körpern. Z Für Anal
647 Chem 22:366–382. https://doi.org/10.1007/BF01338151

Królczyk JB, Dawidziuk T, Janiszewska-Turak E, Sołowiej B (2016) Use of Whey and Whey Preparations in the Food Industry – a Review. Pol J Food Nutr Sci 66:157–165. https://doi.org/10.1515/pjfns-2015-0052

Lamberti C, Cristina L, Acquadro E, et al (2016) Validation of a mass spectrometry-based method for milk traces detection in baked food. Food Chem 199:119–127. https://doi.org/10.1016/j.foodchem.2015.11.130

Layman DK, Lönnerdal B, Fernstrom JD (2018) Applications for α-lactalbumin in human nutrition. Nutr Rev 76:444–460. https://doi.org/10.1093/nutrit/nuy004

Lee J-Y, Kim CJ (2010) Determination of allergenic egg proteins in food by protein-, mass spectrometry-, and DNA-based methods. J AOAC Int 93:462–477

Li S, Wang J, Wei H, et al (2012) Identification of Bovine Casein Phosphorylation Using Titanium Dioxide Enrichment in Combination with Nano Electrospray Ionization Tandem Mass Spectrometry. J Integr Agric 11:439–445. https://doi.org/10.1016/S2095-3119(12)60029-X

Lifschitz C, Szajewska H (2015) Cow's milk allergy: evidence-based diagnosis and management for the practitioner. Eur J Pediatr 174:141–150. https://doi.org/10.1007/s00431-014-2422-3

Liu G, Xiong YL, Butterfield DA (2000) Chemical, Physical, and Gel-forming Properties of Oxidized Myofibrils and Whey- and Soy-protein Isolates. J Food Sci 65:811–818. https://doi.org/10.1111/j.1365-2621.2000.tb13592.x

Losito I, Introna B, Monaci L, et al (2013) Development of a method for the quantification of caseinate traces in Italian commercial white wines based on liquid chromatography-electrospray ionization-ion trap-mass spectrometry. J Agric Food Chem 61:12436–12444. https://doi.org/10.1021/jf4034909

Lutter P, Parisod V, Weymuth H (2011) Development and validation of a method for the quantification of milk proteins in food products based on liquid chromatography with mass spectrometric detection. J AOAC Int 94:1043–1059

MANEA I, AILENEI E, DELEANU D (2016) Overview of food allergy diagnosis. Clujul Med 89:5–10. https://doi.org/10.15386/cjmed-513

Mariotti F, Tomé D, Mirand PP (2008) Converting Nitrogen into Protein—Beyond 6.25 and Jones' Factors. Crit Rev Food Sci Nutr 48:177–184. https://doi.org/10.1080/10408390701279749

Mattarozzi M, Milioli M, Bignardi C, et al (2014) Investigation of different sample pre-treatment routes for liquid chromatography–tandem mass spectrometry detection of caseins and ovalbumin in fortified red wine. Food Control 38:82–87. https://doi.org/10.1016/j.foodcont.2013.10.015

Monaci L, Brohée M, Tregoat V, van Hengel A (2011a) Influence of baking time and matrix effects on the detection of milk allergens in cookie model food system by ELISA. Food Chem 127:669–675. https://doi.org/10.1016/j.foodchem.2010.12.113

Monaci L, Losito I, De Angelis E, et al (2013) Multi-allergen quantification of fining-related egg and milk proteins in white wines by high-resolution mass spectrometry. Rapid Commun Mass Spectrom RCM 27:2009–2018. https://doi.org/10.1002/rcm.6662

687 Monaci L, Losito I, Palmisano F, et al (2011b) Towards the quantification of residual milk allergens in
688 caseinate-fined white wines using HPLC coupled with single-stage Orbitrap mass
689 spectrometry. Food Addit Contam Part Chem Anal Control Expo Risk Assess 28:1304–1314.
690 https://doi.org/10.1080/19440049.2011.593191

691 Monaci L, Losito I, Palmisano F, Visconti A (2010a) Identification of allergenic milk proteins markers in
692 fined white wines by capillary liquid chromatography-electrospray ionization-tandem mass
693 spectrometry. J Chromatogr A 1217:4300–4305.
694 https://doi.org/10.1016/j.chroma.2010.04.035

695 Monaci L, Nørgaard JV, Hengel AJ van (2010b) Feasibility of a capillary LC/ESI-Q-TOF MS method for
696 the detection of milk allergens in an incurred model food matrix. Anal Methods 2:967–972.
697 https://doi.org/10.1039/C0AY00151A

698 Monaci L, Pilolli R, De Angelis E, et al (2014) Multi-allergen detection in food by micro high-
699 performance liquid chromatography coupled to a dual cell linear ion trap mass spectrometry.
700 J Chromatogr A 1358:136–144. https://doi.org/10.1016/j.chroma.2014.06.092

701 Montowska M, Fornal E (2019) Absolute quantification of targeted meat and allergenic protein
702 additive peptide markers in meat products. Food Chem 274:857–864.
703 https://doi.org/10.1016/j.foodchem.2018.08.131

704 Montowska M, Fornal E (2018) Detection of peptide markers of soy, milk and egg white allergenic
705 proteins in poultry products by LC-Q-TOF-MS/MS. LWT 87:310–317.
706 https://doi.org/10.1016/j.lwt.2017.08.091

707 Münch G, Schicktanz D, Behme A, et al (1999) Amino acid specificity of glycation and protein-AGE
708 crosslinking reactivities determined with a dipeptide SPOT library. Nat Biotechnol 17:1006–
709 1010. https://doi.org/10.1038/13704

710 Neta P, Pu Q-L, Kilpatrick L, et al (2007) Dehydration versus deamination of N-terminal glutamine in
711 collision-induced dissociation of protonated peptides. J Am Soc Mass Spectrom 18:27–36.
712 https://doi.org/10.1016/j.jasms.2006.08.016

713 Newsome GA, Scholl PF (2013) Quantification of allergenic bovine milk α(S1)-casein in baked goods
714 using an intact $^{15}$N-labeled protein internal standard. J Agric Food Chem 61:5659–5668.
715 https://doi.org/10.1021/jf3015238

716 Ng-Kwai-Hang KF, Hayes JF, Moxley JE, Monardes HG (1987) Variation in Milk Protein Concentrations
717 Associated with Genetic Polymorphism and Environmental Factors. J Dairy Sci 70:563–570.
718 https://doi.org/10.3168/jds.S0022-0302(87)80042-5

719 Paradela A, Albar JP (2008) Advances in the Analysis of Protein Phosphorylation. J Proteome Res
720 7:1809–1818. https://doi.org/10.1021/pr7006544

721 Parker CH, Khuda SE, Pereira M, et al (2015) Multi-allergen Quantitation and the Impact of Thermal
722 Treatment in Industry-Processed Baked Goods by ELISA and Liquid Chromatography-Tandem
723 Mass Spectrometry. J Agric Food Chem 63:10669–10680.
724 https://doi.org/10.1021/acs.jafc.5b04287

725 Paul M, Somkuti GA (2009) Degradation of milk-based bioactive peptides by yogurt fermentation
726 bacteria*. Lett Appl Microbiol 49:345–350. https://doi.org/10.1111/j.1472-
727 765X.2009.02676.x

Pescuma M, Hébert EM, Mozzi F, Valdez GF de (2007) Hydrolysis of whey proteins by Lactobacillus acidophilus, Streptococcus thermophilus and Lactobacillus delbrueckii ssp. bulgaricus grown in a chemically defined medium. J Appl Microbiol 103:1738–1746. https://doi.org/10.1111/j.1365-2672.2007.03404.x

Pilolli R, De Angelis E, Godula M, et al (2014) Orbitrap$^{TM}$ monostage MS versus hybrid linear ion trap MS: application to multi-allergen screening in wine. J Mass Spectrom JMS 49:1254–1263. https://doi.org/10.1002/jms.3453

Pilolli R, De Angelis E, Monaci L (2017) Streamlining the analytical workflow for multiplex MS/MS allergen detection in processed foods. Food Chem 221:1747–1753. https://doi.org/10.1016/j.foodchem.2016.10.110

Pilolli R, De Angelis E, Monaci L (2018) In house validation of a high resolution mass spectrometry Orbitrap-based method for multiple allergen detection in a processed model food. Anal Bioanal Chem 410:5653–5662. https://doi.org/10.1007/s00216-018-0927-8

Planque M, Arnould T, Delahaut P, et al (2019) Development of a strategy for the quantification of food allergens in several food products by mass spectrometry in a routine laboratory. Food Chem 274:35–45. https://doi.org/10.1016/j.foodchem.2018.08.095

Planque M, Arnould T, Dieu M, et al (2016) Advances in ultra-high performance liquid chromatography coupled to tandem mass spectrometry for sensitive detection of several food allergens in complex and processed foodstuffs. J Chromatogr A 1464:115–123. https://doi.org/10.1016/j.chroma.2016.08.033

Planque M, Arnould T, Dieu M, et al (2017a) Liquid chromatography coupled to tandem mass spectrometry for detecting ten allergens in complex and incurred foodstuffs. J Chromatogr A 1530:138–151. https://doi.org/10.1016/j.chroma.2017.11.039

Planque M, Arnould T, Renard P, et al (2017b) Highlight on Bottlenecks in Food Allergen Analysis: Detection and Quantification by Mass Spectrometry. J AOAC Int 100:1126–1130. https://doi.org/10.5740/jaoacint.17-0005

Popping B, Diaz-Amigo C (2017) European Regulations for Labeling Requirements for Food Allergens and Substances Causing Intolerances: History and Future. J AOAC Int. https://doi.org/10.5740/jaoacint.17-0381

Poulsen NA, Jensen HB, Larsen LB (2016) Factors influencing degree of glycosylation and phosphorylation of caseins in individual cow milk samples. J Dairy Sci 99:3325–3333. https://doi.org/10.3168/jds.2015-10226

Prado M, Ortea I, Vial S, et al (2016) Advanced DNA- and Protein-based Methods for the Detection and Investigation of Food Allergens. Crit Rev Food Sci Nutr 56:2511–2542. https://doi.org/10.1080/10408398.2013.873767

Prasad R, Venkata RSA, Ghokale P, et al (2018) Cow's Milk-related Symptom Score as a predictive tool for cow's milk allergy in Indian children aged 0–24 months. Asia Pac Allergy 8:. https://doi.org/10.5415/apallergy.2018.8.e36

Products NRC (US) C on TO to I the NA of A (1988) Factors Affecting the Composition of Milk from Dairy Cows. National Academies Press (US)

Qi K, Liu T, Yang Y, et al (2019) A rapid immobilized trypsin digestion combined with liquid chromatography – Tandem mass spectrometry for the detection of milk allergens in baked food. Food Control 102:179–187. https://doi.org/10.1016/j.foodcont.2019.03.017

Ree R, Varland S, Arnesen T (2018) Spotlight on protein N-terminal acetylation. Exp Mol Med 50:. https://doi.org/10.1038/s12276-018-0116-z

Reid JR, Coolbear T, Ayers JS, Coolbear KP (1997) The action of chymosin on κ-casein and its macropeptide: Effect of pH and analysis of products of secondary hydrolysis. Int Dairy J 7:559–569. https://doi.org/10.1016/S0958-6946(97)00062-9

Restani P, Gaiaschi A, Plebani A, et al (1999) Cross-reactivity between milk proteins from different animal species. Clin Exp Allergy J Br Soc Allergy Clin Immunol 29:997–1004. https://doi.org/10.1046/j.1365-2222.1999.00563.x

Rodriguez J, Gupta N, Smith RD, Pevzner PA (2008) Does Trypsin Cut Before Proline? J Proteome Res 7:300–305. https://doi.org/10.1021/pr0705035

Sathe SK, Sharma GM (2009) Effects of food processing on food allergens. Mol Nutr Food Res 53:970–978. https://doi.org/10.1002/mnfr.200800194

Schoemaker AA, Sprikkelman AB, Grimshaw KE, et al (2015) Incidence and natural history of challenge-proven cow's milk allergy in European children--EuroPrevall birth cohort. Allergy 70:963–972. https://doi.org/10.1111/all.12630

Sicherer SH, Sampson HA (2014) Food allergy: Epidemiology, pathogenesis, diagnosis, and treatment. J Allergy Clin Immunol 133:291–307; quiz 308. https://doi.org/10.1016/j.jaci.2013.11.020

Slangen CJ, Visser S (1999) Use of mass spectrometry To rapidly characterize the heterogeneity of bovine alpha-lactalbumin. J Agric Food Chem 47:4549–4556. https://doi.org/10.1021/jf990212j

Taylor SL, Baumert JL (2010) Cross-contamination of foods and implications for food allergic patients. Curr Allergy Asthma Rep 10:265–270. https://doi.org/10.1007/s11882-010-0112-4

Taylor SL, Nordlee JA, Niemann LM, Lambrecht DM (2009) Allergen immunoassays—considerations for use of naturally incurred standards. Anal Bioanal Chem 395:83–92. https://doi.org/10.1007/s00216-009-2944-0

Tolin S, Pasini G, Simonato B, et al (2012) Analysis of commercial wines by LC-MS/MS reveals the presence of residual milk and egg white allergens. Food Control 28:321–326. https://doi.org/10.1016/j.foodcont.2012.05.015

Van Vlierberghe K, Gavage M, Dieu M, et al (2019) Selection of universal peptide biomarkers for the detection of the allergen hazelnut in food trough a comprehensive, high resolution mass spectrometric (HRMS) based approach. Food Chem 125679. https://doi.org/10.1016/j.foodchem.2019.125679

Wal J-M (2002) Cow's milk proteins/allergens. Ann Allergy Asthma Immunol 89:3–10. https://doi.org/10.1016/S1081-1206(10)62115-1

Wei Y, Sathe SK, Teuber SS, Roux KH (2003) A Sensitive Sandwich ELISA for the Detection of Trace Amounts of Cashew ( *Anacardium occidentale* L.) Nut in Foods. J Agric Food Chem 51:3215–3221. https://doi.org/10.1021/jf025977x

Westerhout J, Baumert JL, Blom WM, et al (2019) Deriving individual threshold doses from clinical food challenge data for population risk assessment of food allergens. J Allergy Clin Immunol 144:1290–1309. https://doi.org/10.1016/j.jaci.2019.07.046

Zhang J, Lai S, Zhang Y, et al (2012) Multiple reaction monitoring-based determination of bovine α-lactalbumin in infant formulas and whey protein concentrates by ultra-high performance liquid chromatography–tandem mass spectrometry using tryptic signature peptides and synthetic peptide standards. Anal Chim Acta 727:47–53. https://doi.org/10.1016/j.aca.2012.03.034

EUR-Lex - 32011R1169 - EN - EUR-Lex. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32011R1169. Accessed 4 May 2020a

About dairy cows. https://www.ciwf.com/farmed-animals/cows/dairy-cows/. Accessed 12 Nov 2019b

PeptideMass. https://web.expasy.org/peptide_mass/. Accessed 28 Jan 2019c

Critical review on proteotypic peptide marker tracing for six allergenic ingredients in incurred foods by mass spectrometry | Elsevier Enhanced Reader. https://reader.elsevier.com/reader/sd/pii/S0963996919306337?token=BE8B6BF7CF0ACEDED9322B844545BC62DD12471C09450BA43A245266566C0B215D3A1B0C9196744F40159CD14695F576. Accessed 12 Feb 2020d

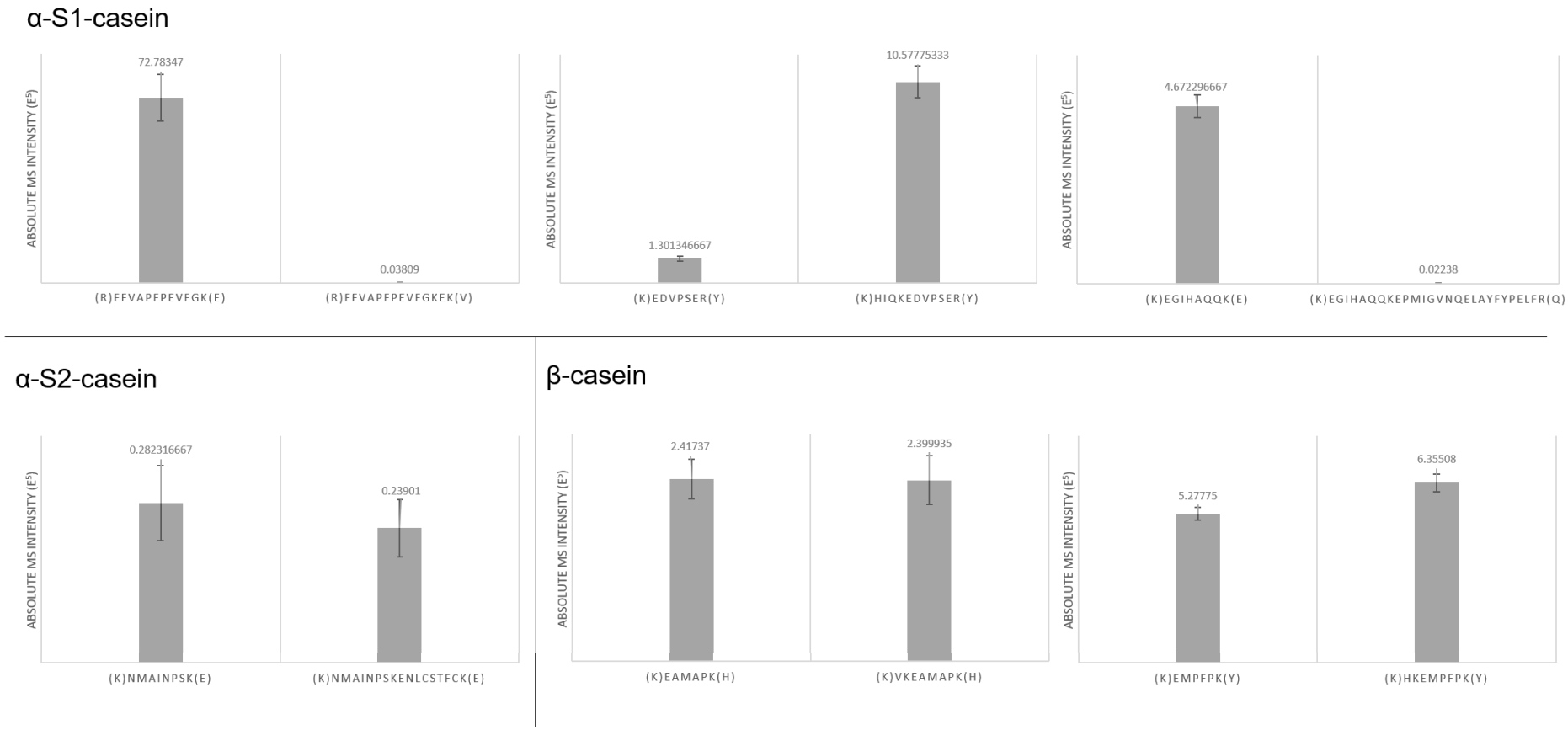



**Fig 1:** Tryptic peptides with observed missed cleavage in a lysine-glutamic acid motif. 6 tryptic peptides from 3 major milk proteins were found to contain missed cleavages. Results are expressed as the means of the peptide absolute MS signal intensity ± 1 SD for 3 biological replicates in raw freeze dried whole milk

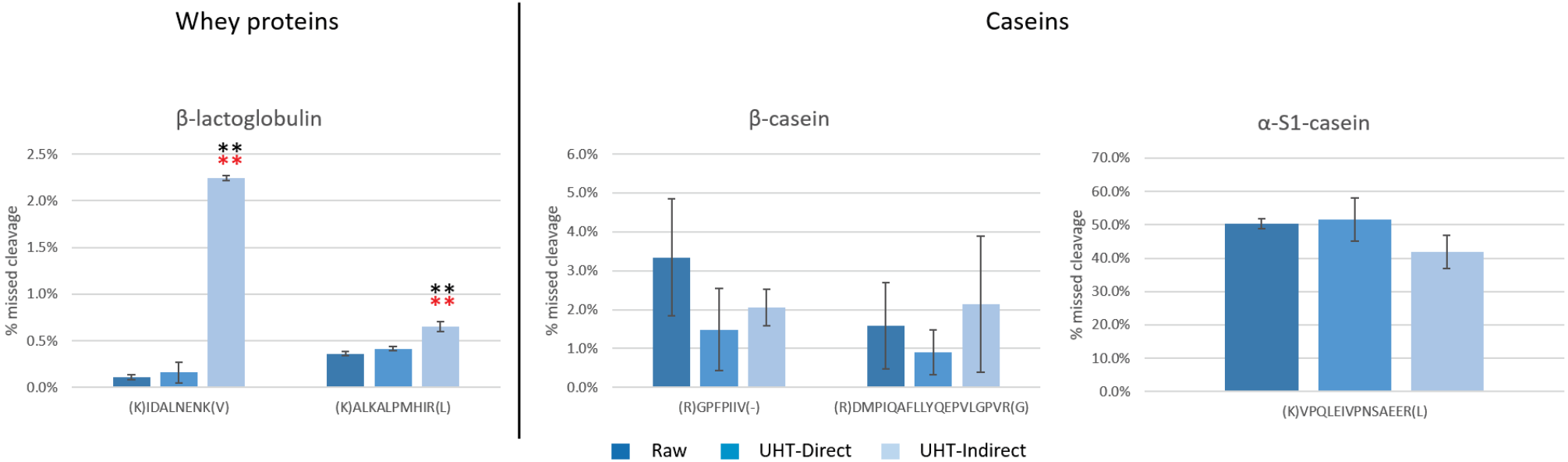



**Fig 2:** Comparison of % missed cleaved peptides in freeze-dried whole milk (unprocessed milk) and freeze-dried UHT milk in an direct (steam injection) and indirect (tubular heat exchange) way. Percentage of missed cleaved peptides was calculated by the ratio the MS signal intensity of the missed cleaved version of the peptide to the total MS signal intensity for that peptide, considering all version of the peptide, expressed as the mean percentage ± 1 SD for 3 biological replicates. Significance levels were calculated by Student's t-test, comparing values from freeze-dried whole milk to freeze-dried UHT milk in an indirect way (black) and freeze-dried UHT milk in an direct way to freeze-dried UHT milk in an indirect way (red) (*=$p<0.05$; **=$p<0.01$, ***=$p<0.001$). Other comparisons showed no significance

.

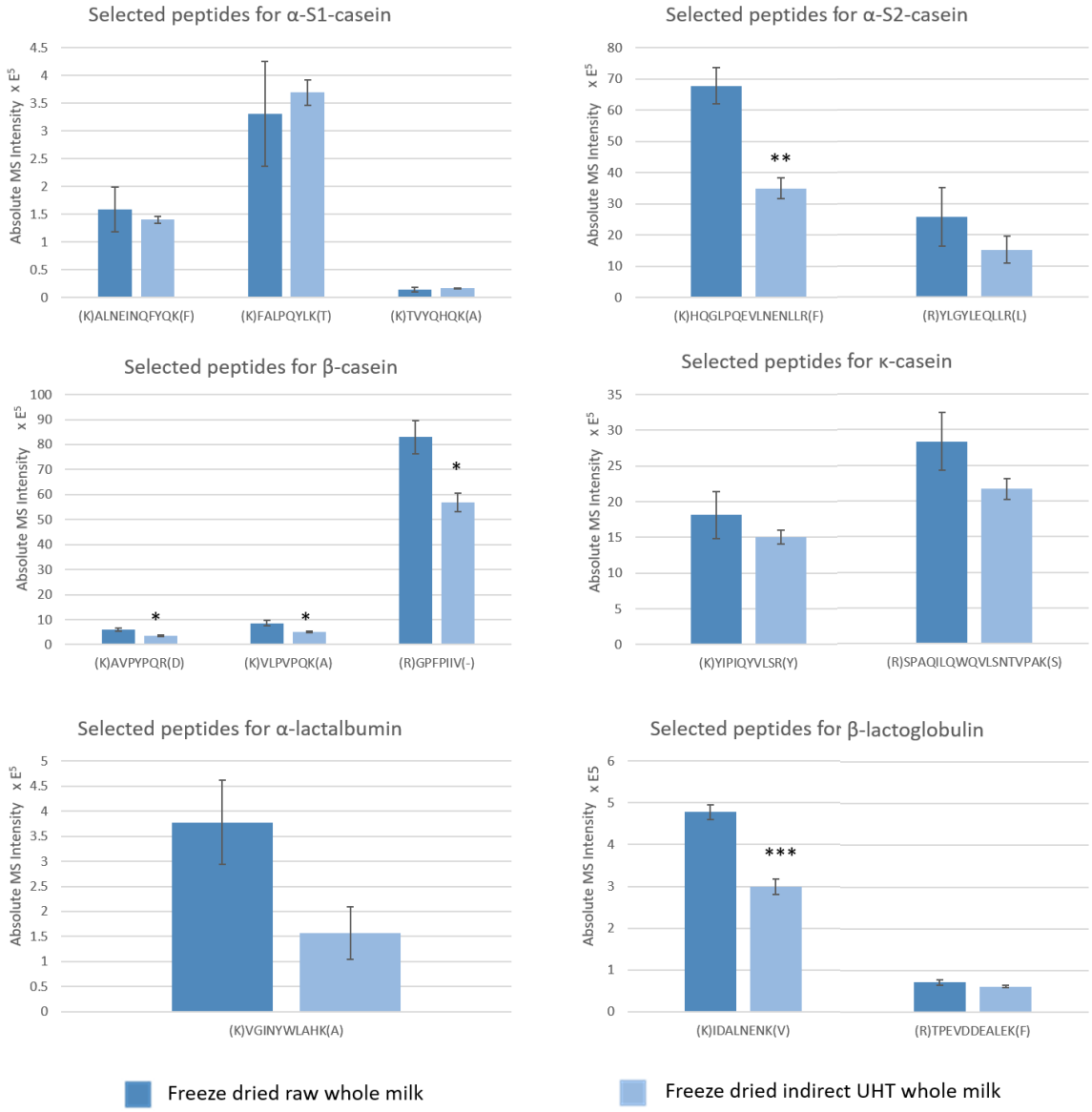



**Fig 3:** Comparison of absolute MS intensities of tryptic peptides that were retained after several selection criteria (table 1) from major milk proteins α-S1-casein, α-S2-casein, β-casein, κ-casein, α-lactalbumin and β-lactoglobulin in freeze-dried whole milk (unprocessed milk) and freeze-dried UHT milk (heat treated and induction of the Maillard reaction). MS signal intensities are expressed as the means of the peptide MS signal intensity ± 1 SD for 3 biological replicates. Significance levels were calculated by Student's t-test, comparing values from freeze-dried whole milk to freeze-dried UHT milk (*=$p<0.05$; **=$p<0.01$, ***=$p<0.001$)

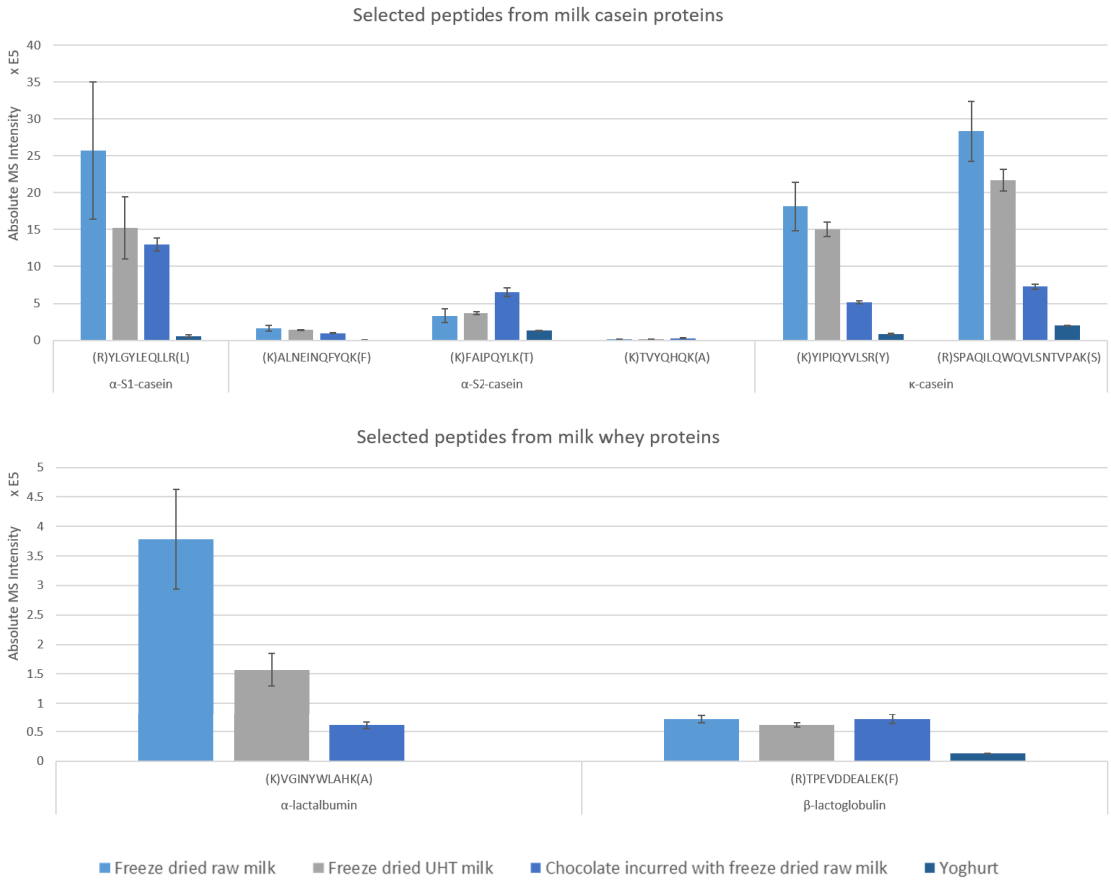



**Fig 4:** Comparison of absolute MS intensities of tryptic peptides that were retained after several selection criteria from the major milk proteins α-S1-casein, α-S2-casein, κ-casein, α-lactalbumin and β-lactoglobulin in freeze-dried whole milk (unprocessed milk), freeze-dried UHT milk (heat treated and induction of the Maillard reaction), chocolate incurred with freeze-dried raw milk (fat-rich environment) and yoghurt (fermented milk, low pH environment). MS signal intensities are expressed as the means of the peptide MS signal intensity ± 1 SD for 3 biological replicates

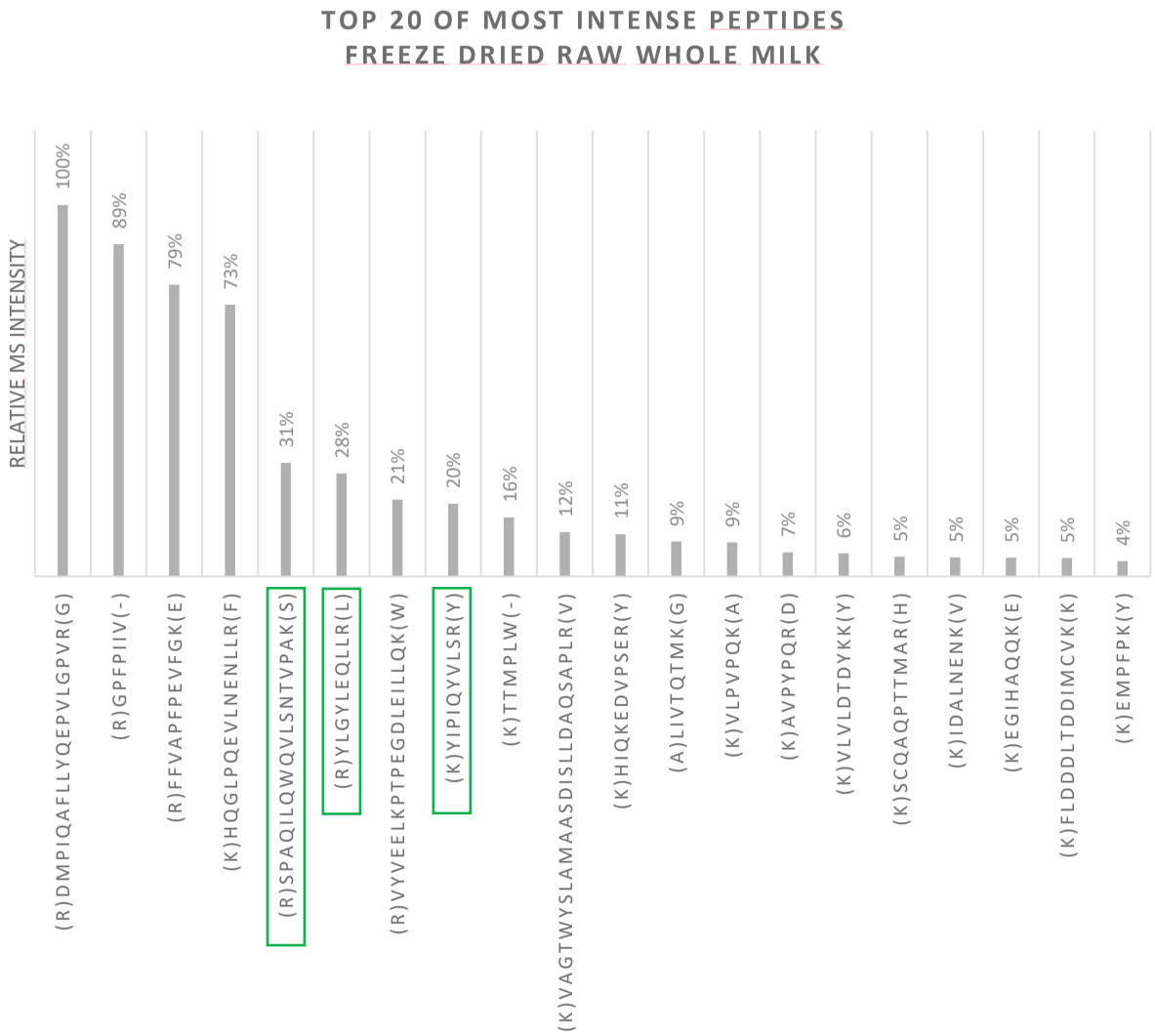



**Fig 5:** Top 20 of most intense detected peptides in freeze-dried raw whole milk. Green-framed peptides are peptides included in our final biomarker selection. Values depicted are the means of relative MS signal intensities ± 1 SD for 3 biological replicates, relative to the most intense peptide.

| Protein | Peptides (>6 AA, <30 AA) | Position | Detected | AA composition | Motifs and sequences | Post-translational modifications |
|---|---|---|---|---|---|---|
| α-S1-casein | FFVAPFPEVFGK | 38-49 | Yes | - | Observed missed cleavage through KE motif | - |
| | **YLGYLEQLLR** | **106-115** | **Yes** | **-** | **-** | **-** |
| | **HQGLPQEVLNENLLR** | **23-37** | **Yes** | **-** | **-** | **-** |
| | VPQLEIVPNSAEER | 121-134 | Yes | - | - | Potential S-phosphorylation |
| | EPMIGVNQELAYFYPELFR | 148-166 | Yes | Methionines | - | - |
| | EGIHAQQK | 140-147 | Yes | - | Observed missed cleavage through KE motif | - |
| | DIGSESTEDQAMEDIK | 58-73 | Yes | Methionines | - | Potential S-phosphorylation |
| | EDVPSER | 99-105 | Yes | - | Observed missed cleavage through KE motif | - |
| | LLILTCLVAVALARPK | 2-18 | No | Cysteines | Part of sequence lies within 20 first amino acids of the protein | - |
| | QMEAESISSSEEIVPNSVEQK | 74-94 | No | Methionines, N-terminal Q | - | Potential S-phosphorylation |
| α-S2- casein | NAVPITPTLNR | 130-140 | Yes | - | KR motif at N-terminus | - |
| | **FALPQYLK** | **189-196** | **Yes** | **-** | **-** | **-** |
| | **ALNEINQFYQK** | **96-106** | **Yes** | **-** | **-** | **-** |
| | ENLCSTFCK | 48-56 | Yes | Cysteines | - | - |
| | TVDMESTEVFTK | 153-164 | Yes | Methionines | Flanked by KK motif at N- and C-terminus | Potential S-phosphorylation |
| | **TVYQHQK** | **197-203** | **Yes** | **-** | **-** | **-** |
| | FPQYLQYLYQGPIVLNPWDQVK | 107-128 | Yes | - | KR motif at C-terminus | - |
| | NMAINPSK | 40-47 | Yes | Methionines | Observed missed cleavage through KE motif | - |
| | NANEEEYSIGSSSEESAEVATEEVK | 61-85 | No | - | - | Potential S-phosphorylation |
| | EQLSTSEENSK | 141-152 | No | - | KK motif at C-terminus | Potential S-phosphorylation |
| | FFIFTCLLAVALAK | 3-16 | No | Cysteines | Part of sequence lies within 20 first amino acids of the protein | - |
| | NTMEHVSSSEESIISQETYK | 17-36 | Yes | Methionines | Part of sequence lies within 20 first amino acids of the protein | Potential S-phosphorylation |
| | AMKPWIQPK | 204-212 | Yes | Methionines | KP motif | - |
| **β-casein** | **GPFPIIV** | **218-224** | **Yes** | **-** | **-** | **-** |

| | | | | | | |
|---|---|---|---|---|---|---|
| | **VLPVPQK** | **185-191** | **Yes** | **-** | **-** | **-** |
| | **AVPYPQR** | **192-198** | **Yes** | **-** | **-** | **-** |
| | DMPIQAFLLYQEPVLGPVR | 199-217 | Yes | Methionines | - | - |
| | FQSEEQQQTEDELQDK | 48-63 | Yes | - | - | Potential S-phosphorylation |
| | VLILACLVALALAR | 3-16 | No | Cysteines | Part of sequence lies within 20 first amino acids of the protein | - |
| | ELEELNVPGEIVESLSSSEESITR | 17-40 | No | - | Part of sequence lies within 20 first amino acids of the protein | - |
| **κ-casein** | **YIPIQYVLSR** | **45-55** | **Yes** | **-** | **-** | **-** |
| | **SPAQILQWQVLSNTVPAK** | **90-107** | **Yes** | **-** | **-** | **-** |
| | SFFLVVTILALTLPFLGAQEQNQEQPIR | 4-31 | No | - | Part of sequence lies within 20 first amino acids of the protein | - |
| | SCQAQPTTMAR | 108-118 | Yes | Methionines, cysteines | - | - |
| | HPHPHLSFMAIPPK | 119-132 | Yes | Methionines | Potential cleavage by chymosin (FM motif) | - |
| **α-lactalbumin** | **VGINYWLAHK** | **118-127** | **Yes** | **-** | **-** | **-** |
| | MMSFVSLLLVGILFHATQAEQLTK | 1-24 | No | Methionines | Part of sequence lies within 20 first amino acids of the protein | - |
| | DDQNPHSSNICNISCDK | 82-98 | Yes | Cysteines | - | Potential N-glycosylayion |
| | FLDDDLTDDIMCVK | 99-112 | Yes | Methionines, cysteines | KK motif at C-terminus | - |
| | LDQWLCEK | 134-141 | Yes | Cysteines | - | - |
| **β-lactoglubulin** | VLVLDTDYK | 108-116 | Yes | - | KK motif at C-terminus | - |
| | **TPEVDDEALEK** | **141-151** | **Yes** | **-** | **-** | **-** |
| | LSFNPTQLEEQCHI | 165-178 | Yes | Cysteines | - | - |
| | VYVEELKPTPEGDLEILLQK | 57-76 | Yes | - | KP motif | - |
| | VAGTWYSLAMAASDISLLDAQSAPLR | 31-56 | Yes | Methionines | - | - |
| | **IDALNENK** | **100-107** | **Yes** | **-** | **-** | **-** |
| | CLLLALALTCGAQALIVTQTMK | 3-24 | No | Methionines, cysteines | Part of sequence lies within 20 first amino acids of the protein | - |
| | WENGECAQK | 77-86 | Yes | Cysteines | KK motif at C-terminus | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| | ALPMHIR | 158-164 | Yes | Methionines | - | - |
| | YLLFCMENSAEPEQSLACQCLVR | 118-140 | Yes | Methionines, cysteines | KK motif at C-terminus | - |

**Table 1:** Biomarker peptide selection for detection of milk protein in food, based on peptide detectability (column 4; yes = detected in all 3 biological replicates in freeze-dried raw whole milk and freeze-dried indirect UHT treated whole milk, No = not detected in all 3 biological replicates), amino acid composition (column 5, exclusion based on presence of certain amino acids such as prone to artificial and post-translational modifications), motifs and sequences (column 6, exclusion based on the presence of certain motifs and sequences that would result in multiple chemical forms of the peptide) and known site-specific post-translational modifications (column 7, potential of multiple variants of the peptide). Peptides in bold fulfill all criteria of potential peptide biomarkers for detection of milk proteins in food. These were used in a further selection based on food processing robustness.

| Protein | Peptide | References |
|---|---|---|
| **α-S1-casein** | (R)YLGYLEQLLR(L)* | (2,12,62–82) |
| **α-S2-casein** | (K)ALNEINQFYQK(F)* | (64,66,68,80,83,84) |
| | (K)FALPQYLK(T)* | (64–66,68,80,84,85) |
| | (K)TVYQHQK(A)* | 85 |
| **κ-casein** | (K)YIPIQYVLSR(Y) | (64,70,80,83,84) |
| | (R)SPAQILQWQVLSNTVPAK(S) | (64,70,80,82) |
| **α-lactalbumin** | (K)VGINYWLAHK(A)* | (62,86) |
| **β-lactoglobulin** | (R)TPEVDDEALEK(F)* | (72,74,75,84,87) |

**Table 2:** Selected peptide biomarkers robust to heat treatment for milk. Peptides with star also show good robustness towards a fat-rich environment.

## 5. Conclusion

The analyte selection is one of the first steps in the development of any analytical method. The proteins of allergenic ingredients are detected and quantified through their constitutive peptides. In this work, the identification of peptide biomarkers was based on an empirical approach with the HRMS analysis of various processed test matrices containing either egg, milk, peanut or hazelnut.

With this approach, hundreds of peptides were identified for each allergenic ingredient tested. These peptides were further filtered based on several criteria to ensure performance of the analytical method. Indeed, a peptide needs to be specific for the allergen of interest, robust to food processing and not subjected to amino acid modifications or missed enzymatic cleavages to be considered as a good biomarker for the detection and quantification of an allergen. Specificities of each allergenic ingredient were thus considered in the peptide selection. As egg yolk and white can be used separately in food industry, peptide biomarkers from both parts have therefore to be selected. Multiple protein isoforms and origin variation issues were also considered for peanut and hazelnut.

Based on HRMS analysis and selection criteria, around ten peptides per allergenic ingredient were identified as potential biomarkers for the development of a quantitative method. These peptides and the related proteins are listed in Table 5.

*Table 5 – List of peptide biomarkers that fulfil the selection criteria (specificity for the allergenic ingredient, robust to food process and not prone to amino acid modifications or missed enzymatic cleavages). The peptides are listed in front of their respective proteins and allergenic ingredient. The protein accession numbers in the UniProt database have been placed in brackets except for Ara h 1 and Ara h 3 as they correspond to multiple isoforms as detailed in the section 4.2 (page 70).*

| | | Protein | Peptide biomarkers |
|---|---|---|---|
| Milk | Casein | Alpha-S1-casein (P02662) | FFVAPFPEVFGK, YLGYLEQLLR, HQGLPQEVLNENLLR, EGIHAQQK |
| | | Alpha-S2-casein (P02663) | NAVPITPTLNR, ALNEINQFYQK, FALPQYLK |
| | | Kappa casein (P02668) | YIPIQYVLSR, SPAQILQWQVLSNTVPAK |
| | Whey | Alpha lactalbumin (P00711) | VGINYWLAHK |
| | | Beta lactoglobulin (P02754) | TPEVDDEALEK, VYVEELKPTPEGDLEILLQK, IDALNENK |
| Egg | White | Lysozyme (P00698) | FESNFNTQATNR |
| | | Ovalbumin (P01012) | GGLEPINFQTAADQAR, HIATNAVLFFGR |
| | | Ovotransferrin (P02789) | SAGWNIPIGTLIHR, GAIEWEGIESGSVEQAVAK, VEDIWSFLSK, FYTVISSLK |
| | Yolk | Apolipoprotein B (F1NV02) | HLFLPSSYK, ASFYGLSHAVTK, ALFDYFGYSHDGK, FALSGIVTPGAK, TEEIPPLIENR |
| | | Apovitellenin 1 (P02659) | NFLINETAR |
| | | Vitellogenin 1 (P87498) | TVIVEAPIHGLK, NVNFDGEILK, ATAVSLLEWQR |
| Peanut | | Ara h 1 (multiple isoforms) | DQSSYLQGFSR, GTGNLELVAVR, GSEEEDITNPINLR, GSEEEGDITNPINLR, NNPFYFPSR |
| | | Ara h 3 (multiple isoforms) | AQSENYEYLAFK, NALFVPHYNTNAHSIIYALR, SQSDNFEYVAFK, SQSENFEYVAFK, TANDLNLLILR, TANELNLLILR VYDEELQEGHVLVVPQNFAVAAK, VYDEELQEGHVLVVPQNFAVAGK, WLGLSAEYGNLYR, FNLAGNHEQEFLR, TVNELDLPILNR |
| Hazelnut | | Cor a 8 (Q9ATH2) | AVNDASR |
| | | Cor a 9 (Q8W1C2) | ADIYTEQVGR, ALPDDVLANAFQISR, LNALEPTNR, HFYLAGNPDDEHQR, TNDNAQISPLAGR, EGLYVPHWNLNAHSVVYAIR |
| | | Cor a 11 (Q8S4P9) | LLSGIENFR, ALSQHEEGPPR, ILQPVSAPGHFEAFYGAGGEDPESFYR |

The identification of potential peptide biomarkers was the first step in the development of a method devoted to the quantitative analysis of allergens in food. This list of potential peptide biomarkers was also refined during the development of the UHPLC-MS/MS method and only the best performers in terms of sensitivity and specificity were kept in the final method to quantify multiple allergens.

**Part II**

**Development, production, characterization and evaluation of the performance of the analytical method developed using stable isotope-labelled internal standards ($^{15}$N concatemer and $^{15}$N β-lactoglobulin)**

## 1. Objectives

As already mentioned, food allergen analysis by mass spectrometry is predominantly performed by specific analysis of peptides obtained by an enzymatic digestion of the proteins of the sample (Monaci *et al*, 2018). Targeted proteomics is often used for absolute peptide quantification in combination with isotope dilution, a technique based on the use of an internal standard corresponding to the stable isotope-labelled version of the analyte (Villanueva *et al*, 2014). Isotope-labelled internal standard is used to correct for variability and matrix effects during the analysis (notably for ion suppression effects) (Villanueva *et al*, 2014). As discussed in the section 2.5 of the introduction, depending on the type of internal standard (a peptide, a protein or a concatemer), matrix effects and analyte loss during sample preparation can also be corrected using these internal standards (Planque *et al*, 2017a).

The strategy implemented in this project is based on the use of isotope-labelled concatemers as internal standards. This technique is known for more than a decade in the proteomic field as applied to human drug metabolizing studies (Russell *et al*, 2013) or as a calibrator of mass spectrometry systems (Chawner *et al*, 2012). To the best of our knowledge, this technique has never been applied in food allergen analysis or, more broadly, in food analysis. The objective of this second task was to design, produce, purify and characterize an isotope-labelled concatemer that can be used as an internal standard for the simultaneous quantitative analysis of four allergenic ingredients (egg, milk, peanut and hazelnut). Performance of this recombinant concatemeric protein was evaluated by the analysis of three uncontaminated food matrices spiked with increasing and well-defined concentrations (2.5 ppm to 50 ppm) of the selected allergen extracts. In addition, the potential benefit/power of the concatemer was compared with five synthetic peptides corresponding to tryptic peptides originating from the four considered allergens as well as with β-lactoglobulin, a bovine milk protein that was $^{15}$N isotopically labelled. We thus also designed, produced, purified and characterized this isotope-labelled protein.

From a practical point of view, the design of the concatemer composed of concatenated peptides from the four allergenic ingredients required the identification of these "final" biomarkers. A relatively long list of potential peptide biomarkers was identified with HRMS analysis (see previous chapter). However, this list needed first to be refined to keep only the best peptides in terms of sensitivity detection and specificity. This task was performed by Kaatje Van Vlierberghe (ILVO) during the development of the UHPLC-MS/MS method and is briefly summarised in the next chapter. At the same time, a proof of concept was established with a series of concatemers, only composed of different selected egg peptides. These concatemers were designed, produced, purified and characterized to identify critical parameters and optimal production conditions in order to obtain a high level of expression with a sufficient labelling yield. A high level of expression is essential to ensure the cost-effectiveness of this strategy when compared to synthetic peptides (100 €/mg on average) while a sufficient labelling yield is necessary to avoid any potential risk of false positive analysis. These aspects and the choice of the labelling strategy, a $^{15}$N uniform labelling, is extensively described in a research article published in *Food Chemistry* (Gavage *et al*, 2020a).

## 2. Materials and methods

### 2.1. Introduction

Isotope-labelled concatemers were produced as a recombinant protein. A DNA sequence coding for a protein assembling the selected peptide biomarkers was designed, cloned in an expression vector (pET-17b) and expressed in transformed *E. coli* host cells. Stable isotope-labelling was achieved thanks to the use of isotope-labelled growth medium (Bioexpress Cell Growth Media (U-15N, 98%). Host cells were grown in this culture medium and used isotope-labelled nutrients to express the concatemer, ensuring its uniform $^{15}N$ isotope labelling. Given the cost of this labelled growth medium (1000 €/L), bacteria cell culture and expression conditions were first optimized in classical unlabelled growth medium.

Preliminary tests were first performed with identified and selected egg peptides. Several DNA sequences were designed and expressed in *E. coli* host cells. The egg peptides used for these preliminary tests are listed in Table 6.

*Table 6 - Egg peptides identified as potential biomarkers and used for preliminary tests of concatemer internal standard production. An identifier (EWx or EYx) is used to simplify concatemer writing. Hydrophobicity of each peptide was evaluated based on their grand average of hydropathy (GRAVY) parameter score.*

| | Protein | Id. | Peptide sequence | GRAVY |
|---|---|---|---|---|
| Egg White | Ovalbumin (P01012) | EW1 | GGLEPINFQTAADQAR | -0.538 |
| | | EW2 | HIATNAVLFFGR | 0.783 |
| | Ovotransferrin (P02789) | EW3 | SAGWNIPIGTLIHR | 0.221 |
| | Lysozyme (P00698) | EW4 | FESNFNTQATNR | -1.400 |
| Egg Yolk | Vitellogenin 1 (P87498) | EY1 | ATAVSLLEWQR | 0.136 |
| | | EY2 | NVNFDGEILK | -0.300 |
| | Apolipoprotein B (F1NV02) | EY3 | HLFLPSSYK | -0.133 |
| | Apovitellenin 1 (P02659) | EY4 | NFLINETAR | -0.311 |

As extensively described in the next chapter, 19 peptide biomarkers were considered in the final UHPLC-MS/MS using an associated concatemer internal standard. These peptides are obtained from the digestion with trypsin of several proteins from the four considered allergenic ingredients (egg, milk, peanut and hazelnut). These 19 peptides are listed in Table 7.

*Table 7 – Peptide biomarkers considered for the final UHPLC-MS/MS method and for the associated concatemer internal standard. An identifier is used to simplify concatemer writing. Hydrophobicity of each peptide was evaluated based on their GRAVY parameter score.*

| | | Protein | Id. | Peptide sequence | GRAVY |
|---|---|---|---|---|---|
| Milk | Casein | Alpha-S1-casein (P02662) | P1 | FFVAPFPEVFGK | 0.867 |
| | | | P2 | HQGLPQEVLNENLLR | -0.753 |
| | | | P3 | YLGYLEQLLR | 0.070 |
| | Whey | Beta lactoglobulin (P02754) | P4 | IDALNENK | -0.975 |
| | | | P5 | TPEVDDEALEK | -1.264 |
| | | | P6 | VYVEELKPTPEGDLEILLQK | -0.315 |
| Egg | White | Ovalbumin (P01012) | P7 | GGLEPINFQTAADQAR | -1.200 |
| | | | P8 | HIATNAVLFFGR | 0.342 |
| | | Ovotransferrin (P02789) | P9 | FYTVISSLK | 0.867 |
| | | | P10 | SAGWNIPIGTLIHR | 0.221 |
| | Yolk | Vitellogenin 1 (P87498) | P11 | NVNFDGEILK | -0.300 |
| | | | P12 | TVIVEAPIHGLK | 0.808 |
| Peanut | | Ara h 1 (multiple isoforms) | P13 | GSEEEDITNPINLR | -1.157 |
| | | | P14 | GSEEEGDITNPINLR | -1.107 |
| | | | P15 | GTGNLELVAVR | 0.436 |
| Hazelnut | | Cor a 9 (Q8W1C2) | P16 | ADIYTEQVGR | -0.690 |
| | | | P17 | ALPDDVLANAFQISR | 0.240 |
| | | | P18 | LNALEPTNR | -0.878 |
| | | | P19 | TNDNAQISPLAGR | -0.777 |

## 2.2. DNA sequence design

DNA sequences based on amino acid sequences were designed with Vector NTI Software (Thermo Fisher Scientific). Visual Gene Developer Software (University of California-Davis) was used for mRNA secondary structure visualisation and optimization. Hydrophobicity of each peptide was evaluated based on their grand average of hydropathy (GRAVY) parameter score (Kyte & Doolittle, 1982). The GRAVY value is calculated by adding the hydropathy value for each residue and dividing the sum by the length of the sequence. Positive values indicate hydrophilicity whereas negative values indicate hydrophobicity.

Designed DNA constructs were composed of several regions as illustrated in Figure 32. The constructs were surrounded by two enzymatic restriction sites (*NdeI* and *XhoI*) to be sub-cloned in the expression vector in a directional-dependent manner. The constructs were composed of:

- A priming sequence to enhance mRNA translation. This priming sequence is composed of 36 nucleotides corresponding to the N-term sequence of glutathione S-transferase (GST) or green fluorescent protein (GFP), depending on the constructs. An arginine codon was added to this priming sequence in order to separate the priming sequence from the N-term peptide of the concatenated peptide biomarker during trypsin digestion.
- The concatenated peptide biomarkers (selected egg peptides for preliminary tests or the 19 peptides biomarkers considered in the UHPLC-MS/MS method).
- The GFP (Green Fluorescent Protein) sequence was inserted in some constructs to easily and rapidly control for protein expression and solubility. The GFP is fluorescent under UV light if the protein is expressed and in a soluble form (Waldo *et al*, 1999). DNA sequence encoding GFP was surrounded/flanked by two EcoRV restrictions sites.

- A poly-histidine tag (Poly-H) for concatemer purification using metal affinity chromatography.

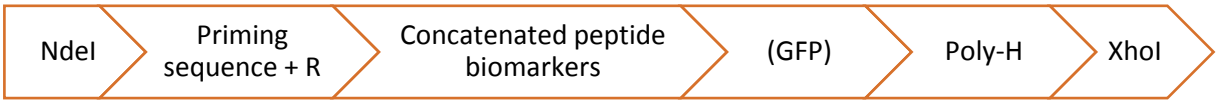



*Figure 32 – Schematic composition of the different designed DNA constructs*

Several important points need to be considered for the design of the concatemer:

1) Peptide hydrophobicity

Hydrophobic and hydrophilic peptides were alternated in the concatemer sequence to obtain a concatemer with a relatively homogenous hydrophobicity and avoid potential high hydrophobicity clusters that could impact concatemer solubility.

2) Translation-associated aspects such as tRNA-mediated codon usage bias and mRNA secondary structures

These aspects are known to impair the translation efficiency (Gorochowski *et al*, 2015). Indeed, a secondary structure with strong interactions as hairpins or loops can have a negative impact on mRNA translation. For a given DNA sequence, the software *Visual Gene Developer* allows to calculate the secondary structure of the associated mRNA. This secondary structure is expressed as a Gibb's free energy (G) profile expressed in kcal/mol/nt (nt=nucleotide). The value of G is inversely proportional to the strength of the interaction (Shabalina *et al*, 2006). Using peptides permutation (conserving hydrophobic/hydrophilic alternation) and silent mutations, the concatemer DNA sequence was optimized to obtain an associated mRNA without intense interaction ($G < -0.2$ kcal/mol/nt). This optimization was applied to "Construct 2-1" and following constructs.

Several concatemers were designed and produced to identify critical parameters and optimal production conditions. The composition and characteristics of the different DNA constructs is summarized in Table 8. The complete nucleotide sequence of each construct is presented in Annex 2 on page 215.

*Table 8 – Composition and characteristics of the different DNA constructs for concatemer production*

| | Priming sequence | Concatenated peptide biomarkers | GFP expression and solubility indicator | Optimized mRNA secondary structure | Expected MW |
|---|---|---|---|---|---|
| **Construct 1-1 (C 1-1)** | Construct not coding for any peptide biomarker | | | | |
| | GST | / | Yes | No | 30 kDa |
| **Construct 1-2 (C 1-2)** | Construct containing one copy of the 8 egg peptide biomarkers | | | | |
| | GST | EW1-EW2-EW4-EW3-EY2-EY1-EY3-EY4 | Yes | No | 41 kDa |
| **Construct 1-3 (C 1-3)** | Construct containing egg white peptide biomarkers in natural abundance | | | | |
| | GST | EW1-EW2-EW4-EW1-EW2-EW4-EW1-EW2-EW3-EW1-EW2-EW1-EW2-EW1-EW2-EW1-EW2-EW1-EW2 | Yes | No | 59 kDa |
| **Construct 1-4 (C 1-4)** | Construct containing egg yolk peptide biomarkers in natural abundance | | | | |
| | GST | EY3-EY2-EY1-EY2-EY1-EY3-EY3-EY4-EY3-EY3 | Yes | No | 42 kDa |
| **Construct 2-1 (C 2-1)** | Construct containing the 8 egg peptide biomarkers with first mRNA secondary structure optimization | | | | |
| | GST | EW3-EW1-EW2-EW4-EY1-EY4-EY3-EY2 | Yes | Yes | 41 kDa |
| **Construct 2-2 (C 2-2)** | Construct similar to Construct 2-1 with N-term ATG methionine codon of GFP expression and solubility indicator replaced by TTA leucine codon | | | | |
| | GST | EW3-EW1-EW2-EW4-EY1-EY4-EY3-EY2 | Yes | Yes | 41 kDa |
| **Construct 2-3 (C 2-3)** | Construct similar to Construct 2-1 with deletion of GFP, expression and solubility indicator sequence | | | | |
| | GST | EW3-EW1-EW2-EW4-EY1-EY4-EY3-EY2 | No | Yes | 14 kDa |
| **Construct 3-1 (C 3-1)** | Construct containing one copy of the 19 final peptide biomarkers | | | | |
| | GST | P13-P4-P1-P5-P2-P8-P16-P18-P3-P7-P12-P14-P11-P10-P6-P17-P9-P15-P19 | No | Yes | 29 kDa |
| **Construct 3-2 (C 3-2)** | Construct containing one copy of the 19 final peptide biomarkers | | | | |
| | GST | P14-P7-P3-P5-P18-P9-P4-P17-P10-P2-P1-P16-P15-P12-P6-P19-P8-P11-P13 | No | Yes | 29 kDa |
| **Construct 3-3 (C 3-3)** | Construct containing one copy of the 19 final peptide biomarkers | | | | |
| | GFP | P11-P6-P8-P7-P2-P10-P19-P14-P3-P18-P9-P13-P15-P12-P17-P16-P1-P4-P5 | No | Yes | 29 kDa |
| **Construct 3-4 (C 3-4)** | Construct containing one copy of the 19 final peptide biomarkers | | | | |
| | GFP | P11-P17-P1-P14-P7-P9-P19-P16-P10-P13-P8-P15-P18-P12-P4-P6-P3-P5-P2 | No | Yes | 29 kDa |

## 2.3. Protein production

Two commercial *E. coli* BL21 (DE3) strains were selected to express the sequence of interest of the different DNA constructs. The first one is One Shot™ BL21(DE3) (Thermo Fischer Scientific) and the second one corresponds to a mutated version of this cell line, One Shot® BL21 Star™ (DE3) (Thermo Fischer Scientific). This second cell line is mutated in the RNaseE gene and is reported to enhance mRNA stability (according to the cell provider (Invitrogen, 2010)).

As extensively described below, bacteria were transformed with the different plasmids by heat shock procedure. Transformed cells were then selected thanks to an ampicillin resistance gene present in the pET-17b expression vector (Figure 33).

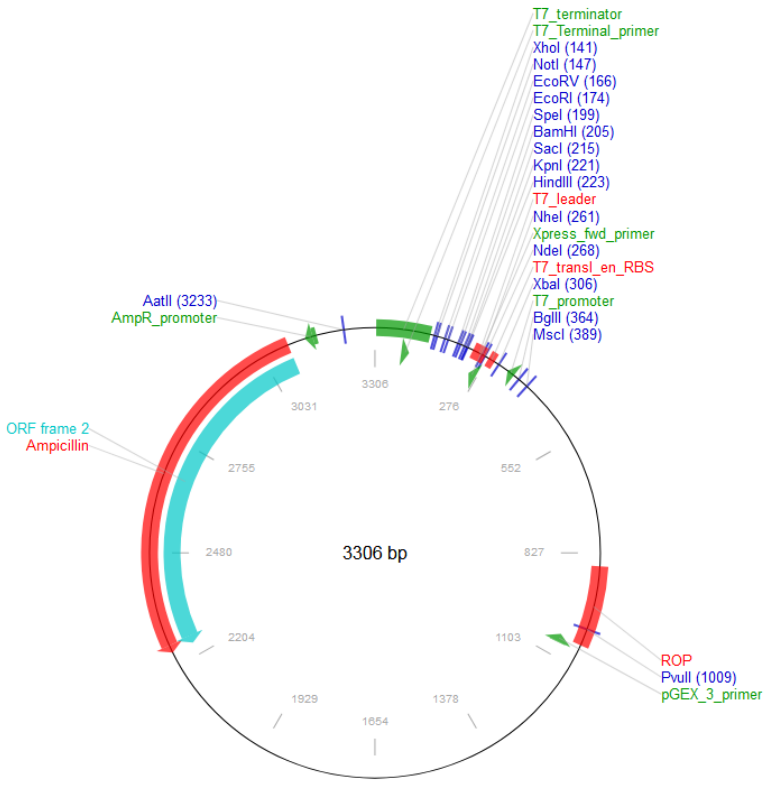



*Figure 33 – Schematic overview of the pE-17b expression vector from Addgene* (from Addgene *website*). *The ampicillin resistance gene and other generic features are highlighted in red, the promoters, primers and terminator in green, the origin of replication in light blue and the restriction sites in dark blue.*

As presented in Figure 34, the identification of critical parameters such as expression yield or protein solubility and the setup of optimal production conditions (bacteria strain, induction time and temperature) were established by comparing two methods of concatemer production.

In the first one, bacteria were seeded on petri dishes to easily observe fluorescence of expressed and soluble GFP. In the second approach, cells were grown in batch mode and the abundance of expressed protein (in total fraction (TF) and soluble fraction (SF)) was evaluated by both SDS-PAGE and Western-Blot (WB) analysis.

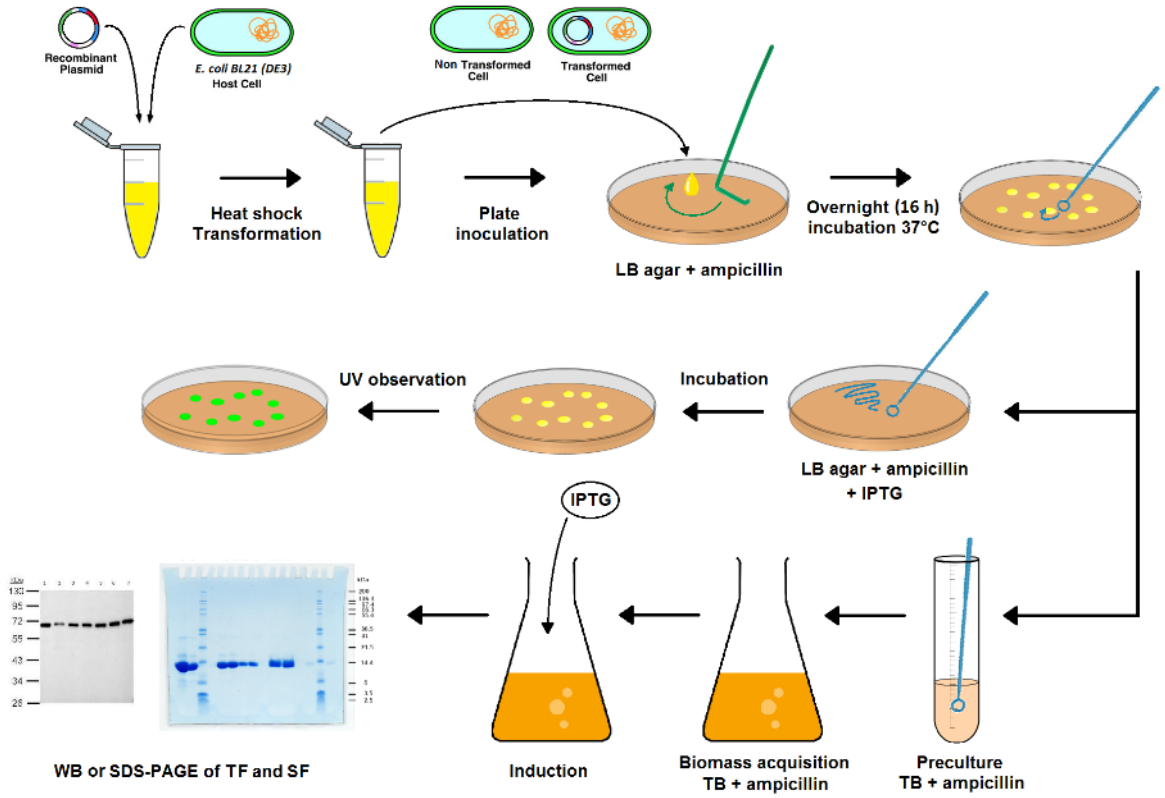



*Figure 34 – Bacteria transformation, cell culture using solid and liquid media. Protein expression analysis was performed directly on cells cultured in petri dishes, by monitoring the GFP associated fluorescence or by SDS-PAGE and WB analyses from cell lysates prepared from bacteria grown in batch mode.*

Below, are the detailed protocols of the different steps in recombinant protein/concatemer production:

a) Bacteria transformation

- 4 µl of pET-17b plasmid (0.1 µg/µl) are added to 100 µl of host cells (One Shot™ BL21(DE3) or One Shot™ BL21 Star™ (DE3)).
- Incubation for 10 min on ice
- Heat-shock of 1 min at 42°C
- Addition of 500 µl of S.O.C. medium (Thermo Fischer Scientific)
- Incubation during 1 h at 37°C under 250 rpm orbital agitation
- Spreading of 100 µl of culture on a petri dish filled with 20 ml of lysogeny broth (LB) medium with agar containing 100 µg/ml ampicillin (Sigma-Aldrich)
- Incubation for 16 h at 37°C incubator

b) Petri dish culture for GFP fluorescence observation

- A petri dish filled with 20 ml of LB agar containing 100 µg/ml ampicillin is spread with 100 µl of 0.2 mM IPTG (Isopropyl β-D-1-thiogalactopyranoside) to induce protein expression
- The petri dish is divided in 5 circular sectors
- Transformed colonies are collected from the petri dish used for cell transformation and spread in each circular sector
- Incubation for 16 h in 30°C or 37°C incubator

c) Batch mode culture: 100 ml

- A colony is collected on the petri dish and seeded in 5 ml of LB medium containing 100 µg/ml ampicillin
- Incubation for 16 h at 37°C under 300 rpm orbital agitation
- Transfer of 2 ml of this first culture in 100 ml of terrific broth (TB) medium (Sigma-Aldrich) containing 100 µg/ml ampicillin placed in a 500 ml Erlenmeyer flask
- Incubation at 37°C under 300 rpm orbital agitation with regular growth control of light absorbance at 660 nm
- When the absorbance between 0.6 and 0.8 is reached, IPTG is added to the medium at a 1 mM final concentration
- Incubation during 4 h at 30° or 37° C or for 16 h at 18°C under 300 rpm orbital agitation
- Harvest 5 ml of cell culture for analysis

d) Batch mode culture: 1000 ml

- A colony is collected on the petri dish and seeded in 5 ml of LB medium containing 100 µg/ml ampicillin
- Incubation for 8 h at 37°C under 300 rpm orbital agitation
- Transfer of 500 µl of this first culture in 25 ml LB medium containing 100 µg/ml ampicillin placed in a 100 ml Erlenmeyer flask
- Incubation for 16 h at 37°C under 300 rpm orbital agitation
- Transfer of 10 ml of this second culture in 500 ml of TB medium containing 100 µg/ml ampicillin placed in a 2000 ml Erlenmeyer flask (operation performed in duplicate)
- Incubation at 37°C under 300 rpm orbital agitation with regular growth control of light absorbance at 660 nm
- When the absorbance between 0.6 and 0.8 is reached, IPTG is added to the medium at a 1 mM final concentration
- Incubation during 4 h at 30° or 37° C or for 16 h at 18°C under 300 rpm orbital agitation
- Harvest 5 ml of cell culture for analysis
- Centrifugation (5000 g for 10 min) and conservation of the pellet at -70°C until protein purification

e) Batch mode culture 1000 ml with $^{15}N$ labelled medium

- Labelled cell growth medium corresponds to “Bioexpress Cell Growth Media (U-15N, 98 %) (10x concentrate)” from Cambridge Isotope Laboratories (CGM-1000-N)
- The 100 ml concentrate labelled medium is first diluted with 900 ml autoclaved water
- A colony is collected on the petri dish and seeded in 5 ml of labelled medium containing 100 µg/ml ampicillin
- Incubation for 8 h at 37 °C under 300 rpm orbital agitation
- Transfer of 250 µl of this first labelled culture in 25 ml labelled medium containing 100 µg/ml ampicillin placed in a 100 ml Erlenmeyer
- Incubation for 16 h at 37°C under 300 rpm orbital agitation
- Transfer of 5 ml of this second labelled culture in 485 ml labelled medium containing 100 µg/ml ampicillin placed in a 2000 ml Erlenmeyer (operation performed in duplicate)
- Incubation at 37°C under 300 rpm orbital agitation with regular growth control of light absorbance at 660 nm
- When the absorbance between 0.6 and 0.8 is reached, IPTG is added to the medium at a 1 mM final concentration
- Incubation for 4 h at 37° or 16 h at 25°C under 300 rpm orbital agitation
- Harvest 5 ml of cell culture for analysis
- Centrifugation (5000 g for 10 min) and conservation of the pellet at -70°C until protein purification

## 2.4. Protein expression analysis

The analysis of the recombinant protein production was performed using 3 techniques.

First, for the petri dish culture, the evaluation of the production was done by the observation of the GFP fluorescence under UV light. An observed fluorescence indicates the presence of the protein and its expression in a soluble form. The absence of fluorescence indicates the absence of the protein or its presence under an insoluble form (inclusion bodies).

For batch culture and control during protein purification, the proteins were detected using Western-blot or SDS-PAGE followed by a Coomassie blue staining.

At the end of the induction period, a 5 ml aliquot of the culture was collected and its optical density (OD) at 660 nm was measured in a spectrophotometer (Ultrospec 10, Biochrom). The 5 ml aliquot was centrifuged (5000 g for 10 min in an Allegra X-12R centrifuge, Beckman Coulter) and the supernatant was discarded. The pellet was next resuspended with 20 mM Tris buffer pH 8, in a volume corresponding to 0.25 ml multiplied by the measured 660 nm absorbance to normalize the different samples based on cell density. Cells were disrupted using a Vibra-Cell™ (Sonics) ultrasonic probe. A volume corresponding to 50 µl of the cell lysates was collected and mixed with 50 µl of SDS-PAGE loading buffer (20 mM Tris pH 6.8, 10 % glycerol, 5 % β-mercaptoethanol, 1 % SDS, 0.01 % bromophenol blue). The first aliquots were removed and corresponded to the total fraction (TF). The remaining cell lysate was next centrifuged (14 800 g for 5 min in a Sorvall™ Legend™ Micro 21R Microcentrifuge, Thermo Scientific). A 50 µl aliquot of the supernatant was collected and mixed with 50 µl SDS-PAGE loading buffer. The second aliquots correspond to the soluble fraction (SF).

For protein purification control, a 50 µl sample was mixed with 50 µl of SDS-PAGE loading buffer if the sample did not contain guanidine hydrochloride (GuHCl), which is not compatible with SDS. If the sample contains GuHCl, proteins from a 50 µl aliquot were first precipitated using a mix of 300 µl water, 400 µl methanol and 100 µl chloroform and then solubilised with 100 µl of SDS-PAGE loading buffer.

A volume of 10 µl of the different samples was loaded in 15-well Mini-PROTEAN® TGX™Resolving Gel (Bio-Rad), with a well dedicated to Color Prestained Protein Standard (New England Biolabs). A constant voltage of 160 V was applied for migration for 45 min.

Depending on the technique, separated proteins were stained with a Coomassie blue solution or transferred to a polyvinylidene fluoride (PVDF) membrane (Amersham™ Hybond™ -P Membranes, GE Healthcare) to be specifically detected with an appropriate IgG antibody. Given the presence of a Poly-H tag of the designed concatemers, a mouse alkaline phosphatase-conjugated anti Poly-H tag IgG antibody was used (ref. ABIN1112523).

### 2.5. Protein purification

Concatemers were purified by metal affinity chromatography thanks to their Poly-H tag. Depending on concatemer solubility, a prior solubilisation step was necessary.

a) Sample preparation for purification

When the concatemer was expressed in inclusion bodies, the following protocol was applied to solubilise the protein:

- Wash buffer 1: 20 mM Tris, 0.5 % Triton X-100, 1 mM EDTA (ethylenediaminetetraacetic acid), pH 8 + 1 mM PMSF (phenylmethylsulfonyl fluoride, a serine protease inhibitor)
- Wash buffer 2: 20 mM Tris, 1 % Triton X-100, 1 mM EDTA, pH 8 + 1 mM PMSF
- Solubilisation buffer: 50 mM Tris, 6 M GuHCl, pH 8

The cell pellet obtained after centrifugation of the 1000 ml batch culture was resuspended in 25 ml of Wash buffer 1. Cells were disrupted using a Vibra-CellTM (Sonics) ultrasonic probe and then centrifuged (20 000 g for 20 min at 4°C in a Sorvall LYNX 6000 Superspeed Centrifuge, Thermo Scientific). The supernatant was discarded and the pellet was resuspended in 15 ml of Wash buffer 2 and incubated for 30 min under soft agitation. The suspension was then centrifuged (20 000 g during 20 min at 4°C). The supernatant was discarded and the pellet was resuspended in 15 ml of water and directly centrifuged (20 000 g during 20 min at 4°C). This water wash was performed twice to remove Triton X-100 and EDTA from the sample. The pellets were then resuspended in 40 ml of Solubilisation buffer and incubated for 16 h under soft agitation at room temperature. Proteins in inclusion bodies were solubilized in GuHCl. The samples were centrifuged (20 000 g during 20 min at 4°C). The supernatant was collected and filtered through 0.2 µm syringe filters (Acrodisc® Syringe Filters with Supor® Membrane, Pall Corporation). Just before affinity purification, imidazole was added to the samples at a 10 mM final concentration. The presence of imidazole allows to reduce unspecific binding during affinity purification (Bornhorst & Falke, 2000).

When the concatemer was expressed in a soluble form, the following protocol was applied:

The cell pellets obtained after centrifugation of the 1000 ml batch culture were resuspended in 40 ml of lysis buffer (50 mM Tris pH 8, 10 mM imidazole, 1 mM PMSF). Cells were disrupted using a Vibra-CellTM (Sonics) ultrasonic probe and then centrifuged twice (20 000 g during 20 min at 4°C). The supernatants were collected and filtered through 0.2 µm syringe filters.

b) Protein metal affinity purification

For the purification of concatemers expressed in inclusion bodies, 6 M GuHCl was added to the different buffers. The protein solution was loaded on Ni Sepharose 6 Fast Flow column (GE Healthcare) equilibrated with lysis buffer (50 mM Tris pH 8, 10 mM imidazole). A washing step was done with in the presence of 20 mM imidazole (in 50 mM Tris pH 8) and the Poly-H labelled concatemer was finally eluted by using a linear imidazole gradient ranging from 20 mM to 250 mM (in 50 mM Tris pH 8). Elution fractions of 1 ml were collected in separated containers. The positive fractions (UV detector) were finally pooled and dialysed against the storage buffer (50 mM Tris pH 8) to eliminate imidazole.

### 2.6. Purified protein characterization for concentration and purity

Protein concentration and purity of purified concatemers were evaluated. Protein purity was assessed using SDS-PAGE with Coomassie blue staining and the gel was analysed with the ImageJ software assuming that all proteins are stained equally well. After background subtraction, protein purity corresponded to the proportion of the density obtained for the band of the protein of interest compared to the density of all the bands in the whole respective lane.

Protein concentration (C expressed in mg/ml) was determined based on its absorbance at 280 nm ($Abs_{280}$), its molecular weight ($MW$ expressed in Da), its purity ($P$) and its molar absorption coefficient at 280 nm ($\varepsilon_{280}$) assuming that molar absorption coefficient at 280 nm of contaminant proteins is similar to that of the protein of interest. Protein concentration was calculated with Equation 1.

$$C = \frac{Abs_{280} * MW * P}{\varepsilon_{280}}$$

*Equation 1 – Protein concentration calculation (expressed in mg/ml) based on its absorbance at 280 nm ($Abs_{280}$), its molecular weight ($MW$ expressed in Da), its purity ($P$) and its molar absorption coefficient at 280 nm ($\varepsilon_{280}$)*

The molar absorption coefficient at 280 nm was calculated based on the protein sequence and more specifically on the corresponding number of tryptophan ($n_W$), tyrosine ($n_Y$) and cysteine ($n_C$) in this sequence (Pace *et al*, 1995). Molar absorption coefficient at 280 nm was calculated with Equation 2.

$$\varepsilon_{280} = 5500\, n_W + 1490\, n_Y + 125\, n_C$$

*Equation 2- Molar absorption coefficient at 280 nm calculation based on protein sequence*

### 2.7. Purified concatemer characterisation: protein sequence and isotopic enrichment

Characterization of the purified concatemers was performed by ultra-high performance liquid chromatography–high resolution mass spectrometry (UHPLC-HRMS) using an Acquity liquid chromatograph equipped with a C18 Acquity BEH130 column (2.1 x 150 mm; 1.7 µm) and coupled to a Xevo G2-XS QTof quadrupole time-of-flight system (Waters). Protein sequence and $^{15}N$ stable isotope enrichment were evaluated with the analysis of their constitutive tryptic peptides. Concatemers were first diluted to 0.1 mg/mL with 50 mM TEAB pH 9.2, to a final volume of 20 µl before being proteolytically digested by adding 0.1 µg of trypsin gold (trypsin:protein ratio of 1:20). Digestion was conducted for 1 h at 37 °C under 300 rpm orbital agitation and stopped by the addition of 1 % (final concentration) of formic acid (FA) followed by a centrifugation (20 000 x g for 5 min). Samples were ten-fold diluted with 5 % acetonitrile (ACN) before UHPLC-HRMS analysis. Peptides (5 µl of sample was injected) were first separated by reverse-phase liquid chromatography using a 20 min water/ACN + 0.1 % FA linear gradient from 5 to 40 % of ACN. Data were acquired in $MS^E$ mode with 0.3 s scan time within the 50 to 2000 m/z mass range. The data were processed using UNIFI software (Waters, Milford, MA, USA) and peptide mapping analysis type with traditional tryptic cleavage rules and setting with $^{15}N$ isotope labelling as a fixed modification when required.

Considering the most intense charge state, the isotopic enrichment or isotope incorporation rate was evaluated for each identified tryptic peptide by comparing the intensity of the peak corresponding to the fully $^{15}N$ labelled (U-$^{15}N$) peptide with other peaks corresponding to partially $^{15}N$ labelled peptides. Practicality (see Figure 35), a $^{13}C$ natural abundance of 1.1 % was considered and hydrogen and oxygen isotopic distributions were neglected in the calculations. Furthermore, only peaks corresponding to peptide with one $^{14}N$ isotope were considered in the calculations. The proportion of U-$^{15}N$ peptide was then calculated by comparing the intensity of the peak corresponding to the (U-$^{15}N$ and U-$^{12}C$) peptide with the peak corresponding to the [(U-1)-$^{15}N$ and U-$^{12}C$] peptide. Protein isotopic enrichment was then evaluated with the exponential trend given by the proportion of the U-$^{15}N$ version of each peptide considering its nitrogen content.

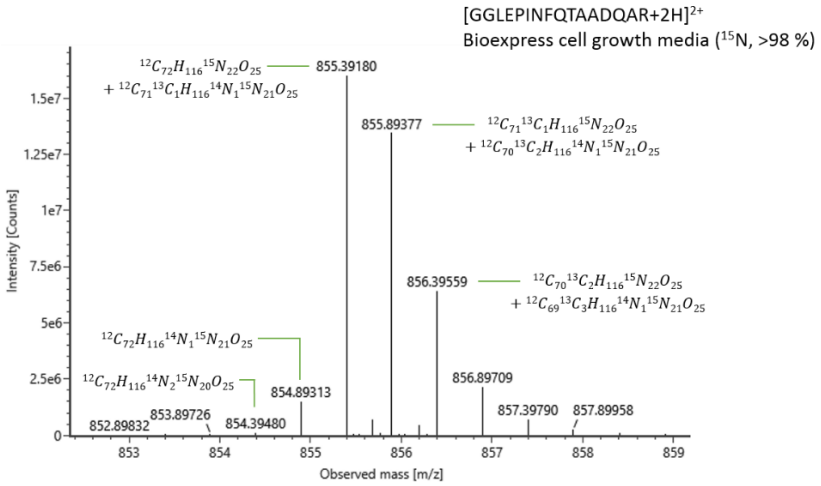



*Figure 35 - Isotope distribution of the doubly charged ion of the peptide GGLEPINFQTAADQAR (from egg white ovalbumin), one of the peptides of a $^{15}N$ isotopically labelled concatemer. Ignoring oxygen and hydrogen isotopes and peptides with more than two $^{14}N$ isotopes, peaks are annotated with their corresponding peptide formulas.*

## 3. Identification of critical parameters and optimal production conditions

Three series of DNA constructs (Construct 1-X, Construct 2-X and Construct 3-X, see Table 8, page 131) were designed and synthetized in order to identify important parameters to consider and optimal production conditions. This optimization was necessary to obtain an isotope-labelled concatemer with sufficient cost-effective production yield and isotopic enrichment. GFP fluorescence observation, SDS-PAGE and WB were used to evaluate the effect on protein expression of the different parameters such as bacterial strain or induction temperature tested.

### 3.1. Preliminary expression assay of constructs encoding egg peptides

A first series of four constructs (Construct 1-1 to Construct 1-4) were designed and cloned into the pET17b expression vector. One Shot™ BL21(DE3) *E. coli* bacteria were transformed with one of these four constructs and different expression conditions were investigated. Bacteria were grown at either 30 or 37°C on solid culture medium (Figure 36) and in batch mode culture medium (Figure 37) with the IPTG addition for 16 h when cells were incubated at 18 °C or for 4 h when cells were grown at 30 or 37°C.

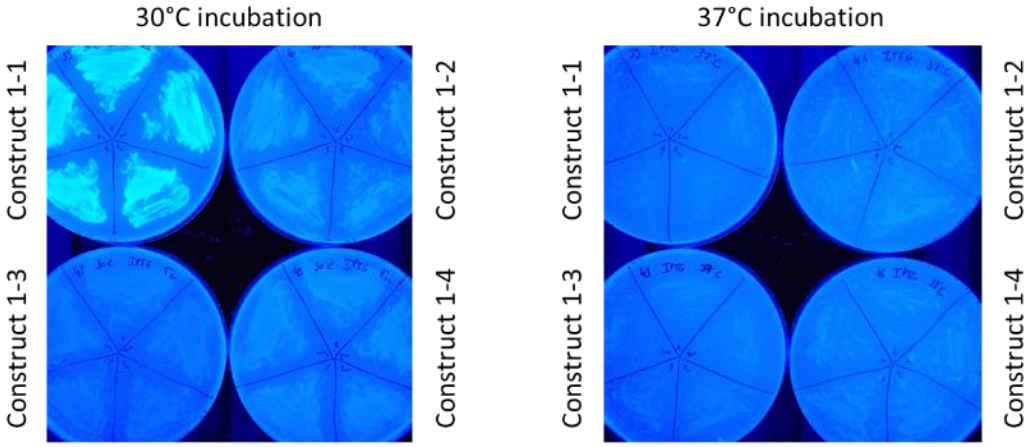



*Figure 36 – One Shot™ BL21(DE3) E. coli cells were transformed with the first series of constructs (Construct 1-1 to Construct 1-4) and then seeded on Petri dishes containing LB agar with 100 µg/ml ampicillin and spread with 100 µl of 0.2 mM IPTG (to induce protein expression) and then incubated at either 30 or 37°C for 16 h. Plates were then illuminated by UV and fluorescence was visualised.*

As observed in Figure 36, a stronger fluorescence is observed, for the four constructs, in bacteria grown at 30 °C when compared to growth performed at 37 °C. This result indicates a higher production yield and/or more soluble protein when cells are exposed to 30 °C. A higher temperature is indeed known to negatively affect the solubility of the expressed proteins accompanied by an increase in the inclusion body formation (Schein & Noteborn, 1988). A difference in the fluorescence intensity was also observed between the different constructs, with the highest fluorescence for "Construct 1-1" (coding only for GFP and not for egg peptides). However, some fluorescence can also be observed in the bacteria transformed with one of the other three constructs, indicating a relative lower but detectable protein expression level.

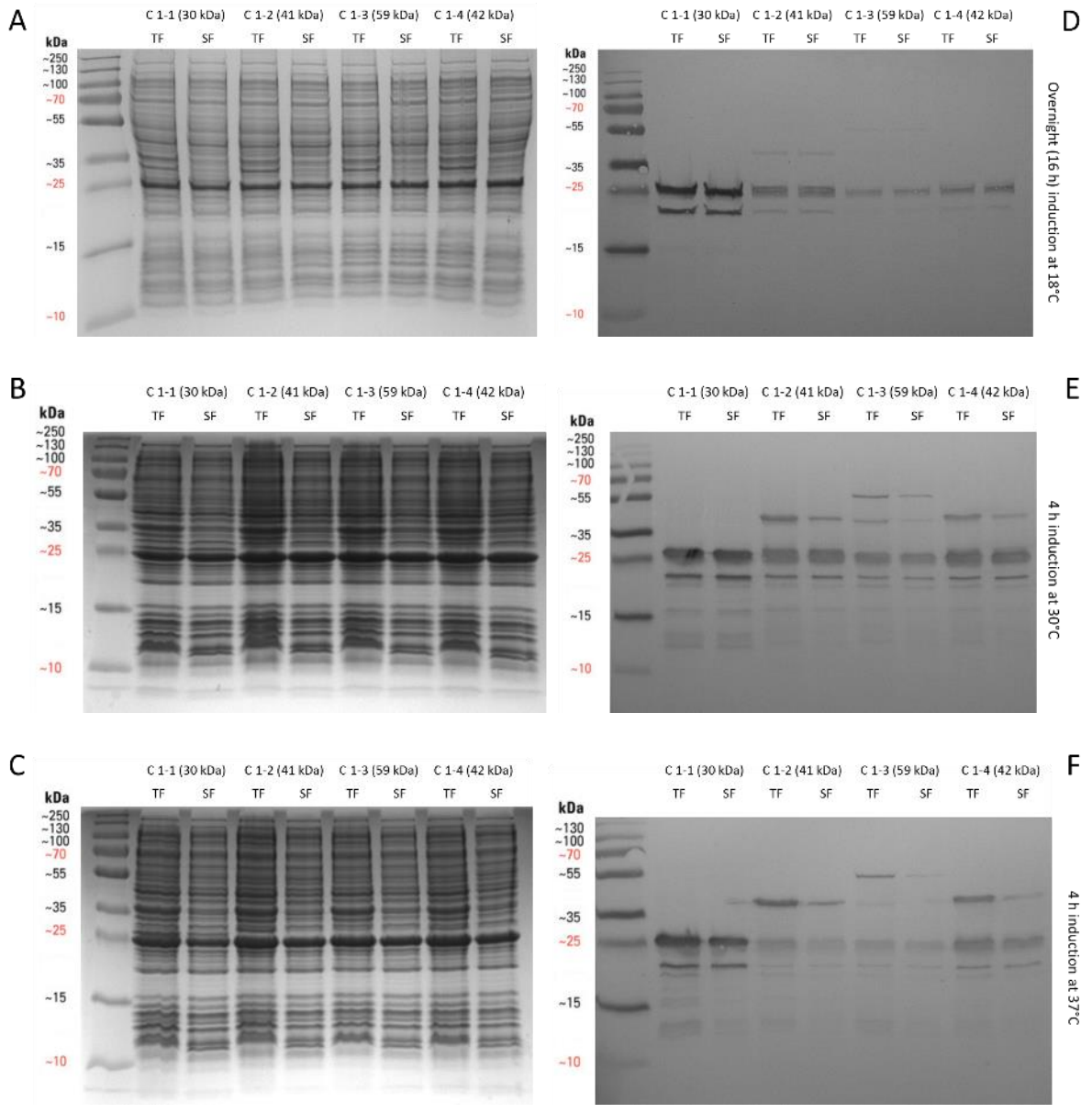



*Figure 37 – One Shot™ BL21 (DE3) E. coli cells were transformed with the first series of constructs (C 1-1 to C 1-4) and cultured in batch mode in 100 ml of TB containing 100 µg/ml ampicillin. Protein expression was induced with the addition of IPTG at a 1 mM final concentration, either for 16 h at 18°C (A and D) or for 4 h at 30°C (B and E) or 37°C (C and F). Harvested cells were analysed by SDS-PAGE with Coomassie blue staining (A, B and C) and by WB with an anti Poly-H antibody (D, E and F). Expected concatemer molecular weight is indicated in brackets. (TF = total fraction, SF = soluble fraction)*

A SDS-PAGE analysis is not specific for the targeted concatemers. Indeed, several proteins present in the total fraction (TF) and the soluble fraction (SF) were detected. No protein corresponding to the expected concatemer molecular weight was observed as an over-expressed protein (Figure 37 A, B, C). The only putative band that could correspond to an over-expressed protein was observed for "Construct 1-1 (C1-1)". This thin band was located between 25 and 35 kDa molecular weight markers, just above the larger band of ~25 kDa observed in every lane. This large band was supposed to be the beta-lactamase, an enzyme produced by *E coli* for ampicillin resistance. This C1-1 specific band was observed in both TF and SF as well as for both temperatures tested.

WB analysis was then used to specifically detect the Poly-H tag of the concatemers (Figure 37 D, E, F). In cell lysates prepared from bacteria transformed with the different constructs, a signal was observed at the expected molecular mass confirming the expression of the different concatemers. However, the expression yield was construct-dependent with a higher amount of expressed protein observed for the C1-1 construct. The protein expression level was also influenced by the induction temperature with a lower expression level for C1-2 to C1-4 at 18 °C. In addition, as expected, bacteria cultures induced at lower temperature produced more soluble proteins. These observations were consistent with the data obtained when GFP fluorescence was analysed (Figure 36).

On the different WB, signals were observed at the expected concatemer molecular weights. However, several other bands were also observed at lower molecular weights. These multiple bands could correspond to concatemer degradation products that kept the Poly-H sequence or the presence of undesired/multiple translation initiation sites. This issue was covered and discussed in the section 3.3.

### 3.2. Optimization of mRNA secondary structure and influence of *E. coli* strain

We have seen that the expression of concatemers was successfully achieved with the first series of constructs tested as a proof of concept. However, the expression level was insufficient to obtain a cost-effective concatemer production. Moreover, expression yield was expected to be even lower with the use of isotope-labelled growth medium, which is less rich in nutrients when compared to classical TB culture medium (Berthold *et al*, 2011).

Constructs containing multiple copies of the same egg peptide, used as a proof of concept, were then discarded and we next switched and focused our efforts on the conditions to optimize expression yield. The parameter tested was the effect of the mRNA secondary structure as it is known to affect the transcript translation efficiency (Gaspar *et al*, 2013).

Strong interactions between nucleotides can lead to the formation of secondary structure features such as hairpins or loops. These structures impair the translation process rate, obstructing ribosome movement along the mRNA (Gorochowski *et al*, 2015) (Figure 38).

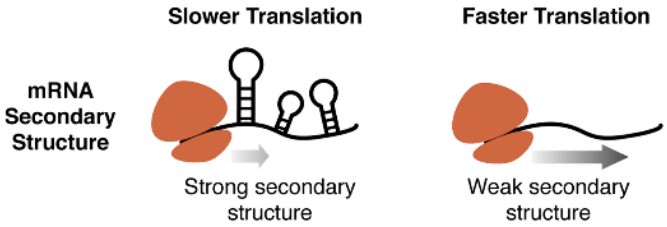



*Figure 38 – Illustration of the negative impact of mRNA secondary structure on translation process* (Gorochowski *et al*, 2015)

The design of the Construct 2-1 was based on the Construct 1-2 content (DNA sequences encoding GST primer, 8 concatenated egg peptides, GFP and poly-H tag). Using egg peptides permutation and silent mutations introductions in the whole construct sequence (using Visual Gene Developer Software), the DNA sequence was modified to optimize mRNA secondary structure and eliminate strong interactions

between nucleotides, preventing hairpin-loop formation, characterized by high negative G values (Figure 39).

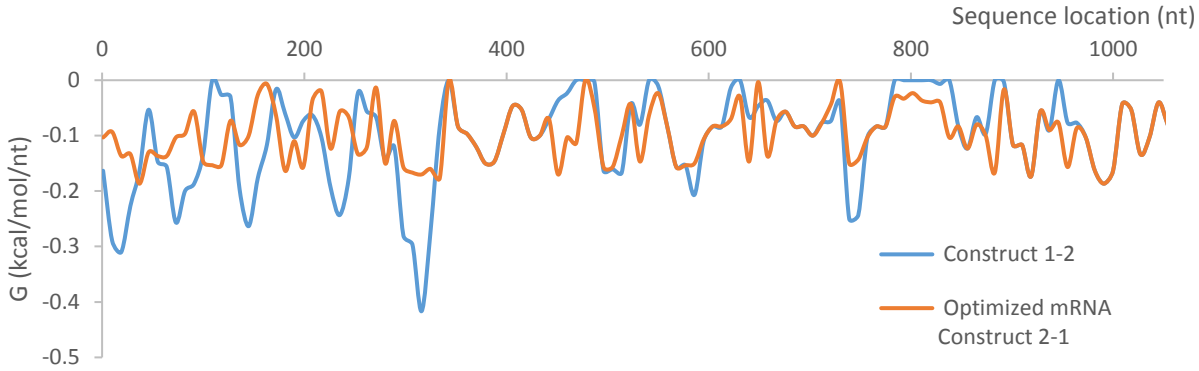



*Figure 39 - DNA sequence optimization of Construct 1-2 to reduce interactions between nucleotides that favour secondary structures in the corresponding mRNA sequence.*

However, a too weak mRNA secondary structure could also lead to rapid mRNA degradation by RNases (such as RNase E) (Del Campo *et al*, 2015). To prevent this phenomenon, the use of a second *E. coli* strain was considered. One Shot® BL21 Star™ (DE3) is mutated in the RNaseE gene of the One Shot™ BL21(DE3) stain, conferring enhanced mRNA stability in this strain.

The effect of changes in the mRNA secondary structure and *E. coli* strain was evaluated by the expression level of the concatemer analysed by either GFP expression in bacteria grown on solid culture media (Petri dishes) (Figure 40) or the abundance of the chimeric protein analysed on cell lysates prepared from bacteria cultured in batch mode (Figure 41).

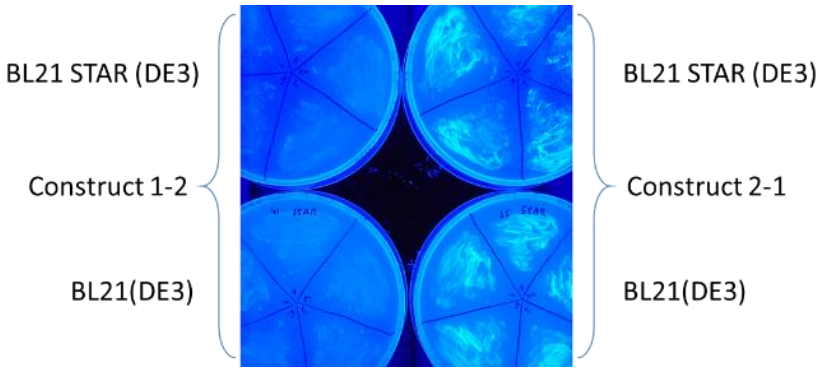



*Figure 40 – Analysis of the expression level of concatemer protein produced by BL21(DE3) bacteria or BL21 Star™ (DE3) transformed with either "Construct 1-2" or "Construct 2-1" and then grown in Petri dishes at 37°C in order to evaluate the effects of E. coli strain and mRNA secondary structure on the expression of the chimeric recombinant protein. Plates were then illuminated with UV light to observe the presence of soluble concatemer containing the GFP sequence.*

In Figure 40, one can observe a more intense fluorescence intensity for the bacteria transformed with the "Construct 2-1" indicating that the mRNA secondary structure could affect/improve protein expression and/or solubility (right part of the Figure). However, the *E. coli* strain does not seem to have any effect on the expression of the protein (left part of the Figure).

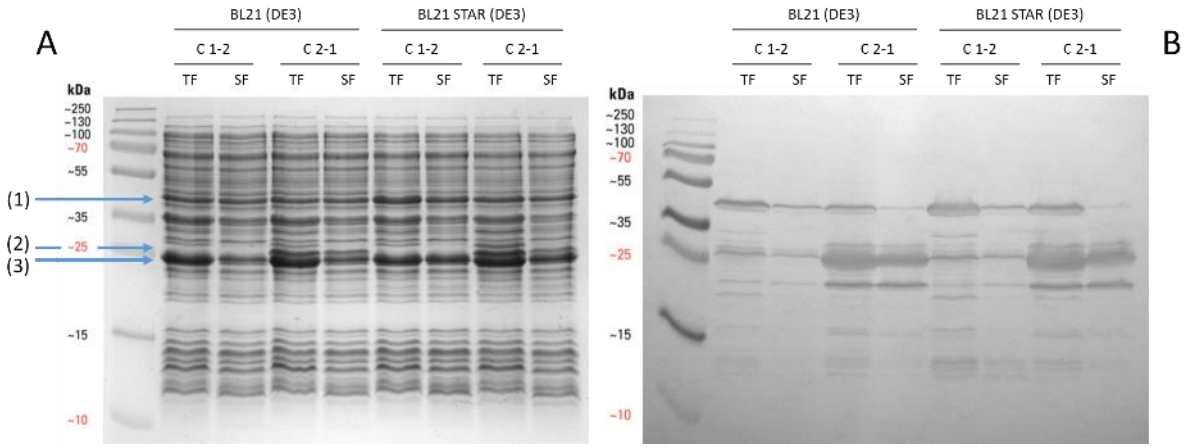



*Figure 41 Effect of the bacterial strain (BL21 (DE3) or BL21 STAR (DE3)) and DNA sequence affecting mRNA secondary structure (constructs C 1-2 or C 2-1) on the expression of the concatemer as analysed by (A) SDS-PAGE and (B) Western Blot using an anti-Poly-H IgG conjugated to alkaline phosphatase. The bacteria were grown at 37°C and the expression of the chimeric protein was induced in the presence of 1 mM IPTG for 4 h at the same temperature. The arrows indicate the expected molecular weight for the 41 kDa concatemer (1), a 25 kDa band (2) and beta-lactamase (3).*

The SDS-PAGE analysis indicated that no protein was found to be over-expressed at the molecular weight expected for the concatamers (41 kDa, arrow (1) on Figure 41 A). A more intense band is observed between the 35 kDa and 55 kDa molecular weight markers, but it could not be concluded that this band corresponds to the protein of interest. Indeed, this band was observed in both total and soluble fractions (TF and SF, respectively) of BL21(DE3) with comparable intensities, which is in contradiction with the results obtained for the WB analysis (Figure 41 B). This band corresponding to C 1-2 seems to be more intense in TF in BL21 STAR (DE3) cells, when compared to the corresponding SF, which is in agreement with the WB analysis (Figure 41 B). A band of approximately 25 kDa (arrow (2) on Figure 41 A), just above beta-lactamase (arrow (3) on Figure 41 A), was also observed in the C 2-1 lanes.

The WB analysis revealed the presence, in every single condition, of the concatemers at the expected molecular mass. However, the concatemers were mainly found in an insoluble form. The use of BL21 STAR (DE3) seemed to slightly improve the protein expression yield. However, the main observed difference was linked to mRNA secondary structure. The protein corresponding to the 41 kDa band seemed to be slightly less intense when produced by BL21 STAR (DE3) when compared to the synthesis by BL21 (DE3). But the main difference is the strong band around 25 kDa only observed for the bacteria transformed with Construct 2-1. In conclusion, the total protein expression was much higher when mRNA secondary structure is optimized.

In the following experiments, efforts were made to try to understand the origin of this 25 kDa band. The objective was to reduce/eliminate the expression of this protein, corresponding to a truncated concatemer, to potentially increase the level of expression of the 41 kDa protein, the protein of interest.

### 3.3. Undesired translation initiation site elimination

HRMS was used to understand the origin of the lower molecular weight undesired protein in concatemer production. The insoluble fraction (inclusion bodies) of a 100 ml TB batch culture of "Construct 2-1" expressed in BL21 STAR (DE3) with a 4 h induction at 37°C was purified by metal affinity chromatography (as described in section 2.5 of the material and method section, page 136) and the different eluted fractions were analysed by SDS-PAGE and Coomassie blue staining (Figure 42). Two main bands were observed at 25 kDa and 41 kDa.

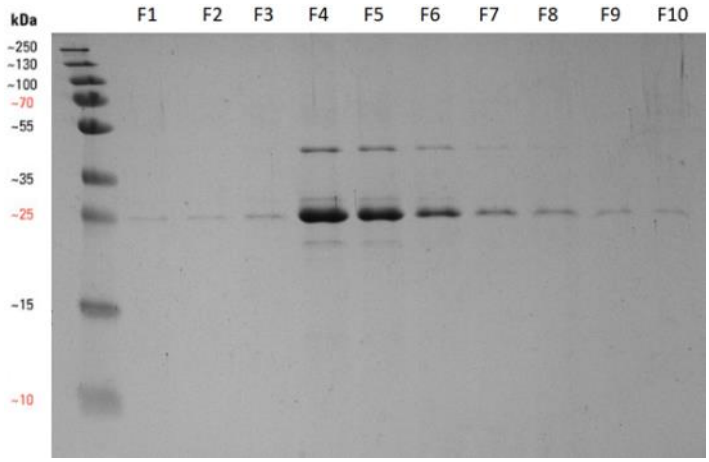



*Figure 42 - SDS-PAGE analysis of eluted fractions (F1 to F10) collected from metal affinity purification of concatemer produced by BL21 STAR (DE3) bacteria cultured in a 100 ml batch at 37 °C. Bacteria were transformed with construct 2-1 and induction of the protein synthesis was triggered by 1 mM IPTG added for 4 h at the same temperature. At the end, the insoluble fraction of the cell lysate was prepared and the chimeric protein was purified by metal affinity chromatography.*

The 25 kDa and 41 kDa bands of the F4 elution fraction were cut out of the gel. The protein of the two bands were separately in-gel digested with trypsin and resulting peptides were analysed by HRMS. This analysis was performed to verify the concatemer sequence and to identify the N-term sequence of the protein of the 25 kDa. Indeed, the C-term domain of the protein was necessarily the Poly-H tag since the purification was based on the affinity of this Poly-H tag for nickel ions of the chromatographic column

As presented in Figure 43, several consecutive steps were implemented to identify the N-term sequence of the protein of the 25 kDa band. The complete concatemer was first *in silico* digested with trypsin cleavage features. Obtained tryptic peptides were sorted by order of appearance in concatemer sequences. Then, HRMS obtained data were treated. Given the difference in band intensity, corresponding to a different protein abundance between the two bands, the signal associated with the different identified peptides was first normalized to obtain comparable data.

The first outcome was the verification of the concatamer sequence. A 100 % sequence coverage was obtained for the analysis of the 41 kDa band (see yellow bar in step 2 in Figure 43). In other words, all tryptic peptides predicted by the *in silico* digestion of the concatemer were observed.

Then, GEELFTGVVPILVELDGDVNGHK was the tryptic peptide observed in the 25 kDa band which was the closest from the N-terminus of the concatemer sequence (framed in red in step 2 in Figure 43). Based on this result, it could be deduced that the N-term sequence of the protein of the 25 kDa band corresponded to a of semi-tryptic fraction of GGSPWGLEPGDIMASK, the tryptic peptide situated at the N-term side of GEELFTGVVPILVELDGDVNGHK in the concatemer sequence. Thus, the m/z of each semi-tryptic peptide of GGSPWGLEPGDIMASK (i.e. SK, ASK, MASK, etc.) was searched in acquired HRMS data (step 3 in Figure 43). A signal corresponding to ASK peptide was observed in the 25 kDa band sample and not in the 41 kDa band sample.

The methionine located at the N-terminus of ASK peptide corresponded to the start codon of GFP in the DNA sequence of "Construct 2-1". An undesired translation initiation site was therefore suspected. Peptide MASK was not observed but this phenomenon could be explained by the fact that the N-term methionine is highly cleaved by the methionine aminopeptidase when followed by an alanine (Frottin *et al*, 2006). Moreover, a sequence (AGGAGAT) close to the Shine-Dalgarno consensus sequence (AGGAGGT) was observed near the methionine ATG codon in the "Construct 2-1" DNA sequence. This Shine-Dalgarno sequence corresponds to a ribosomal binding site in bacteria, necessary for translation initiation (Chen *et al*, 1994).

To validate this hypothesis of an undesired additional translation initiation site, the "Construct 2-1" DNA sequence was modified. The methionine ATG start codon was substituted by a TTA leucine codon. The ATG start codon is an essential element in translation initiation site. If this hypothesis is verified, the ATG start codon mutation would eliminate the undesired translation initiation site and could lead to a higher expression level of the complete concatemer corresponding to 41 kDa. The improvement of concatemer expression yield is the objective of these efforts.

Leucine was selected because it shares comparable physico-chemical properties with methionine. Hence, this mutation would only have a limited impact on concatemer properties. Six codons are corresponding to leucine. The TTA codon was selected because of its low impact on mRNA secondary structure and G profile. The ATG codon was successively *in silico* replaced by each one of the six codons that specify leucine. The mRNA secondary structure associated to these mutated DNA sequences was analysed. As exposed in Figure 44, TTA codon was selected for its low impact on G profile.

Step 1: *In silico* concatemer digestion

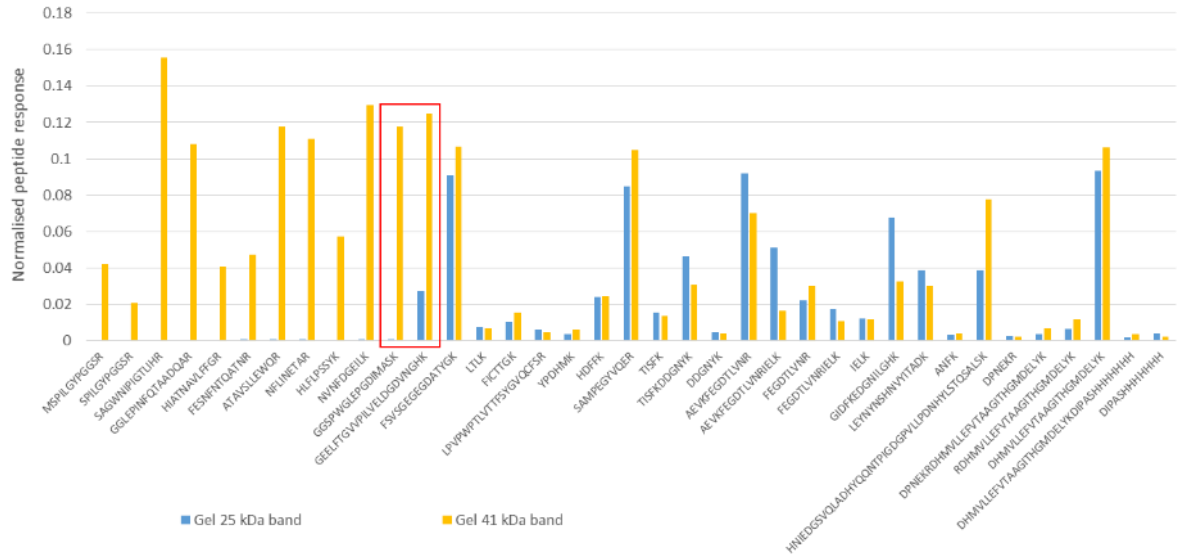
**Full protein sequence**

MSPILGYPGGSRSAGWNIPIGTLIHRGGLEPINFQTAADQARHI
ATNAVLFFGRFESNFNTQATNRATAVSLLEWQRNFLINETARHL
FLPSSYKNVNFDGEILKGGSPWGLEPGDIMASKGEELFTGVVPI
LVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWP
TLVTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFK
DDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNY
NSHNVYITADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPI
GDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITH
GMDELYKDIPASHHHHHHH

**Tryptic peptides**

MSPILGYPGGSR
SAGWNIPIGTLIHR
GGLEPINFQTAADQAR
HIATNAVLFFGR
FESNFNTQATNR
ATAVSLLEWQR
NFLINETAR
HLFLPSSYK
…

Step 2: HRMS analysis of the peptides obtained with in-gel digestion of 25 kDa and 41 kDa bands. Peptides are positioned on the graph, from left to right, by order of appearance in the concatemer sequence. The red box indicates the position of the N-term sequence of the 25 kDa protein in the complete concatemer sequence.




Step 3: Identification of the N-term sequence of the protein of the 25 kDa band

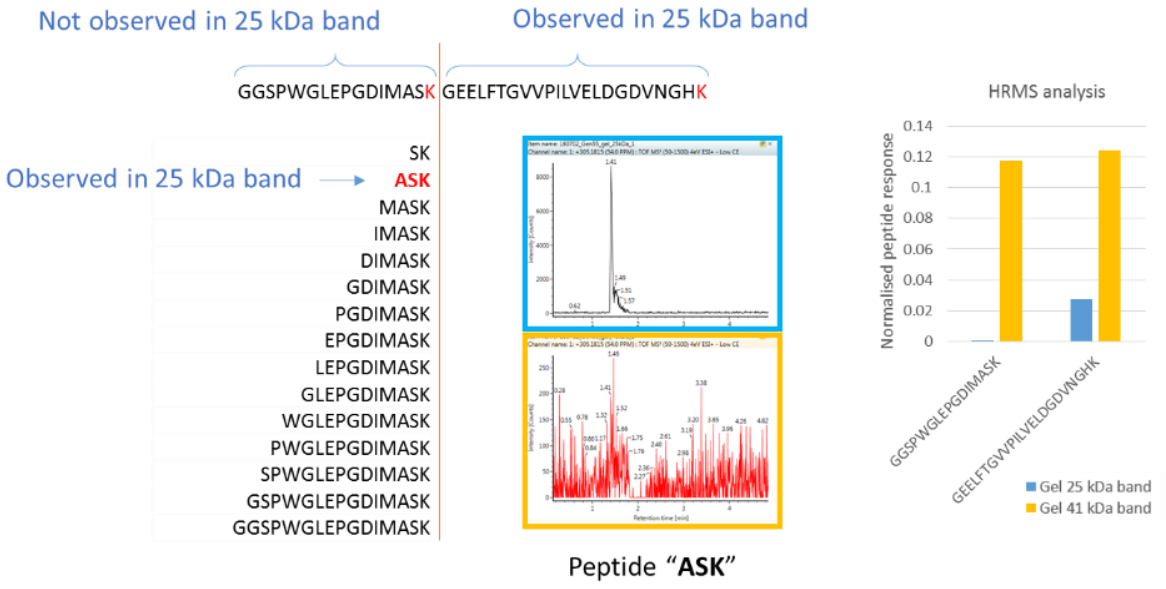



*Figure 43 – Identification of the N-term sequences of the protein of the 25 kDa band by HRMS analysis of peptides obtained by in-gel trypsin digestion*

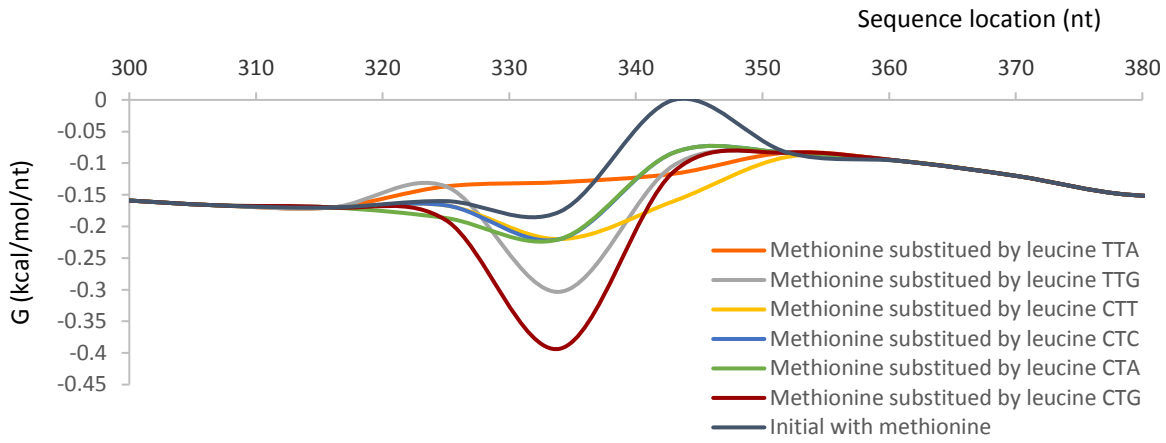



*Figure 44 – Effect of ATG methionine codon substitution on G profile giving information on the mRNA secondary structure.*

A commercial kit (Q5® Site-Directed Mutagenesis Kit, New England Biolabs) was used to proceed to the ATG codon substitution. This kit enables, site-specific mutagenesis of double-stranded plasmid DNA. To do so, specific primers were designed according to kit recommendations (forward primer containing the codon substitution AGGAGATATCTTAGCTAGCAAAGGAGAAGAACTTTTC and reverse primer AGGCACCCCGGATCTTGG). The protocol recommended by the kit manufacturer was followed to induce the mutation. As presented in Figure 45, the plasmid is amplified by PCR using the designed primers (to introduce the ATG codon substitution) and the kit provided DNA polymerase. Then, the use of an enzyme mix, containing a kinase, a ligase and DpnI, allow for circularization of the PCR product and removal of the template DNA. The last step is a high-efficiency transformation into chemically competent cells (provided).

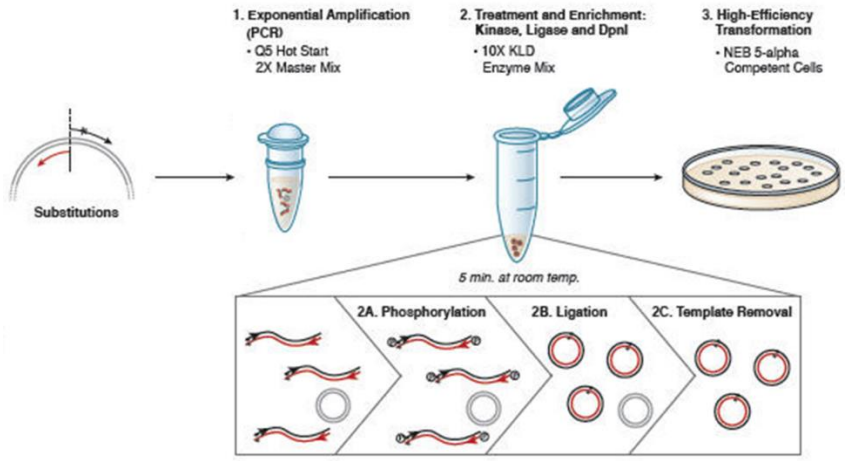



*Figure 45 –Different steps in site directed ATG codon substitution using Q5® Site-Directed Mutagenesis Kit (New England Biolabs). The template plasmid is first amplified by PCR using designed primers (introducing the mutation) and the kit provided DNA polymerase. Then, the use of an enzyme mix, containing a kinase, a ligase and DpnI, allow for circularization of the PCR product and removal of the template DNA. The last step is a high-efficiency transformation into chemically competent cells. Image from* (BioLabs, 2017)

The desired mutation was confirmed by plasmid sequencing by the GIGA Genomics Facility (Figure 46).

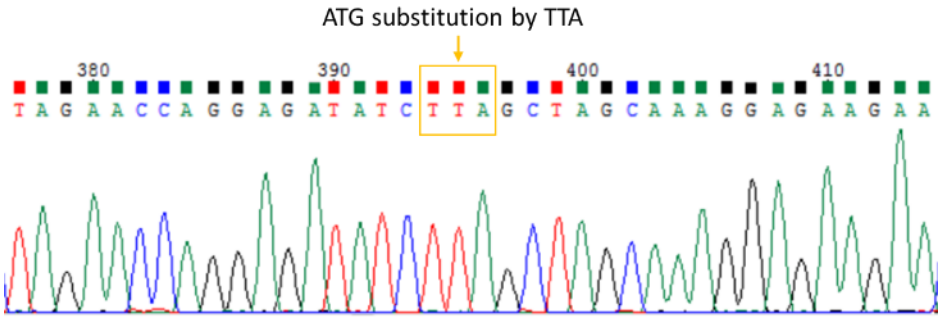



*Figure 46 – Confirmation of ATG start codon substitution by TTA leucine codon by plasmid sequencing*

BL21 Star (DE3) cells were then transformed with the mutated construct (Construct 2-2). Three transformed cell colonies were selected for 100 ml batch mode culture and protein expression (4 h induction with 1 mM IPTG at 37°C). The prepared cell lysates were analysed by anti Poly-H WB (Figure 47). Three clones of the expected mutated construct were selected to maximize the chances to have at least one mutated construct. As observed on the WB analysis, all three selected clones were correctly mutated as no expected 25 kDa protein was produced.

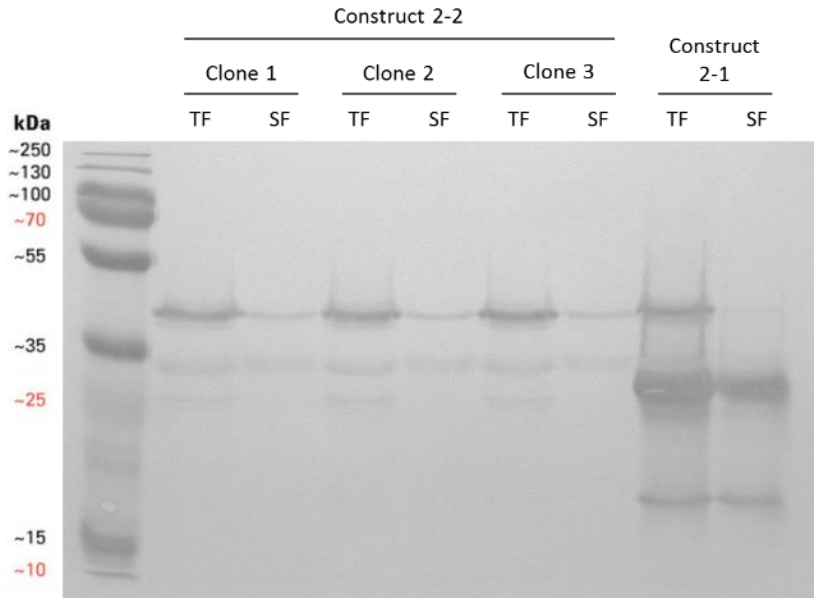



*Figure 47 – Effect of ATG methionine codon substitution by TTA leucine codon in Construct 2-1 on the expression of the concatemer. Western Blot analysis using an anti-Poly-H IgG conjugated to alkaline phosphatase of 100 ml batch mode cultures of BL21 Star (DE3) transformed with three clones of Construct 2-2 (with the substitution) and with Construct 2-1. The bacteria were grown at 37°C and the expression of the chimeric protein was induced in the presence of 1 mM IPTG for 4 h at the same temperature.*

In conclusion, the results obtained confirmed the hypothesis of the presence of an undesired translation initiation site. The 25 kDa band was not observed anymore after the ATG start codon mutation. However, the suppression of this undesired translation initiation site did not allow to improve the expression yield of the complete and expected concatemer corresponding to the 41 kDa band.

### 3.4. Deletion of the GFP sequence

The expression yield of the protein containing the 8 egg peptides was not improved by the elimination of the undesired translation initiation site (Figure 47). We next considered to delete the sequence coding for GFP in the construct. As a reminder, the GFP coding sequence was present in different constructs to easily and rapidly screen and control protein expression and solubility. Once improved production conditions defined, GFP was no longer useful and the sequence was then deleted from the plasmid containing the "Construct 2-1".

Briefly, the plasmid containing the "Construct 2-1" (4298 bp) was digested by EcoRV restriction endonuclease. Two plasmid fragments were then obtained, a first one containing the GFP coding sequence (723 bp) and a larger one containing the rest of the plasmid (3575 bp). This larger fragment was purified, ligated and amplified. This new plasmid with deleted GFP was then called "Construct 2-3". GFP deletion was confirmed with agarose gel electrophoresis analysis (Figure 48). On this gel, three bands were observed in lanes 1 and 3 corresponding to the analysis of circular plasmid. These three bands corresponded to three different forms of the same plasmid (nicked, supercoiled and circular from top to down). These three bands were not observed in lane 2 because the analysed DNA fragment displayed a linear form.

The differences in size between the bands of lane 1 and lane 3 (expected to be 723 bp) confirmed the deletion of the DNA sequence coding for the GFP. The new construct ("Construct 2-3") was therefore coding for a concatemer composed of a GST primer, 8 concatenated egg peptides and a poly-H tag.

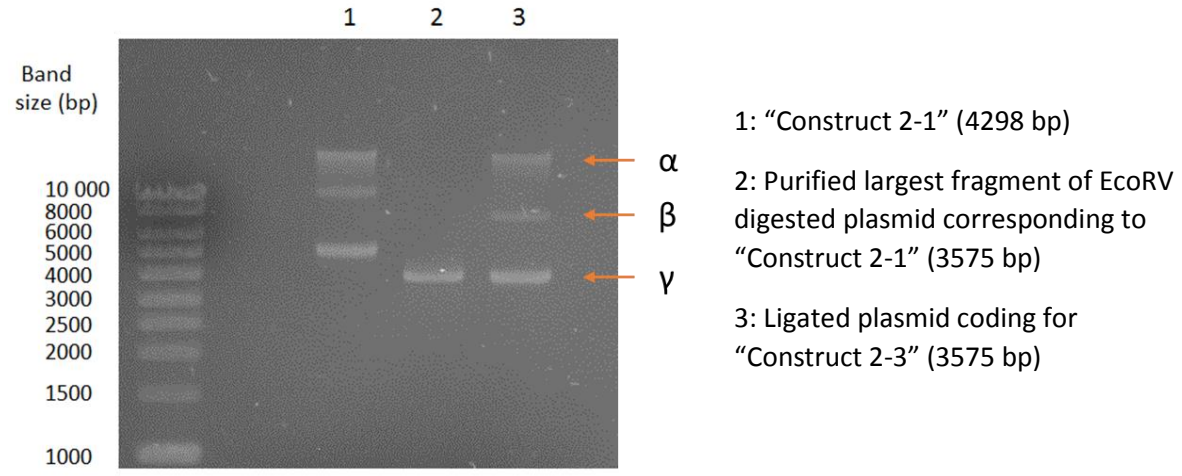



*Figure 48 – 1% agarose gel electrophoresis analysis to confirm GFP sequence deletion. The three bands observed for circular plasmid of lanes 1 and 3 correspond to three forms of the same plasmid (nicked (α), supercoiled (β) and circular (γ)). These three forms were not observed in lane 2 due to the linear form of the plasmid fragment.*

We next tested the expression of the concatemer in bacteria transformed with "Construct 2-3". The effect of *E. coli* strain (BL21(DE3) and BL21 STAR (DE3)) and induction temperature (30° and 37°C) were tested in 100 ml batch mode cultures with a 4 h induction period with 1 mM IPTG.

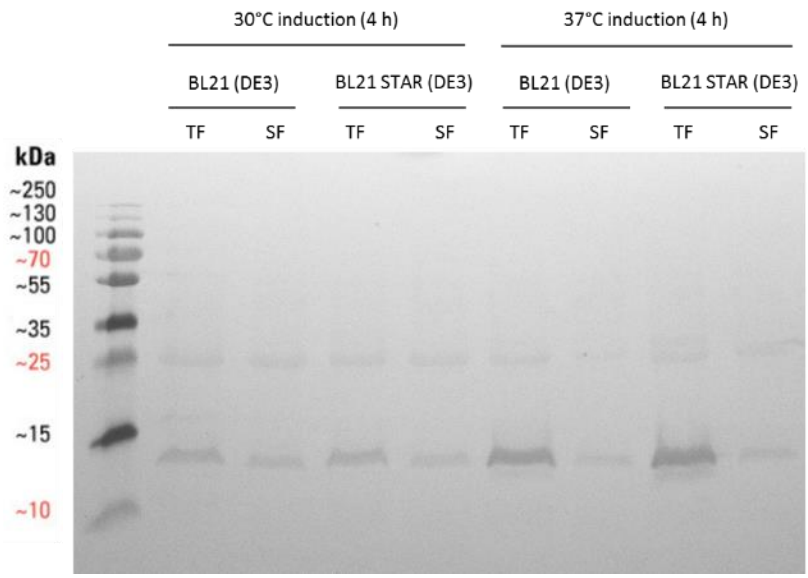



*Figure 49 – Effect of the bacterial strain (BL21 (DE3) and BL21 STAR (DE3)) and induction temperature (30 or 37°C) on the expression of the concatemer. Bacteria were transformed with "Construct 2-3", and cultured in batch mode in 100 ml of TB containing 100 µg/ml ampicillin. Protein expression was induced with the addition of IPTG at a 1 mM final concentration. Harvested cells were analysed by Western Blot using an anti-Poly-H IgG conjugated to alkaline phosphatase.*

The SDS-PAGE analysis of the concatemer production/synthesis in the different batch mode cultures did not allow to highlight the over-expression of the concatemer (data not shown). The WB analysis using the anti-Poly-H IgG was thus used to identify the optimal expression conditions (Figure 49). The concatemer was mainly found in the insoluble form, especially when the expression induction was done at 37 °C. Expression yield was the highest for BL21 STAR (DE3) when the concatemer expression was induced during 4 h at 37 °C. These conditions were selected for the scaling up experiment of this concatemer expression in a 1000 ml-batch production, followed by the purification and characterization of the expressed concatemer.

As the concatemer was mainly found in an insoluble form (Figure 49), the purification protocol optimized for the solubilisation of the proteins in the inclusion bodies was therefore applied. Elution fractions containing the concatemer, identified based on SDS-PAGE analysis, were pooled. Concatemer concentration was measured with 280 nm UV light absorbance and concatemer purity was characterized by SDS-PAGE and gel-image analysis (using ImageJ software as described in section 2.6 of materials and methods, page 136). A total of 3.6 mg of concatemers was produced with a 70 % purity (Figure 50).

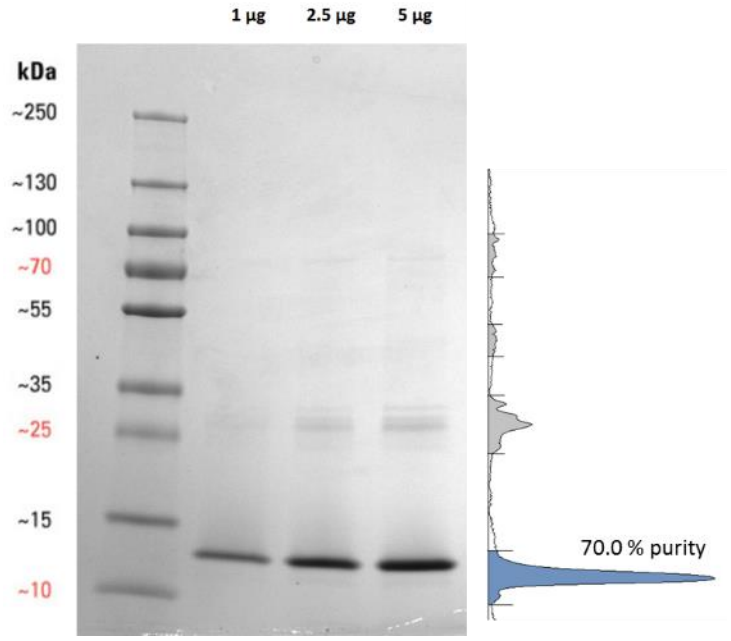



*Figure 50 - SDS-PAGE analysis of purified concatemer encoded by "Construct 2-3". Image analysis of the lane corresponding to 5 µg of proteins was performed with ImageJ software to determine the concatemer purity*

The purified concatemers were digested with trypsin and analysed by mass spectrometry to validate the concatemer sequence. As observed in Figure 51, all 8 egg peptide biomarkers were identified. This last step validated the concatemer sequence and the whole production strategy. The difference in signal intensity associated with the different peptides, theoretically present in equimolar quantities, can be explained by the ionization efficiency which is specific to each peptide. Co-elution of lower intensity peptides with other components can also be part of the explanation.

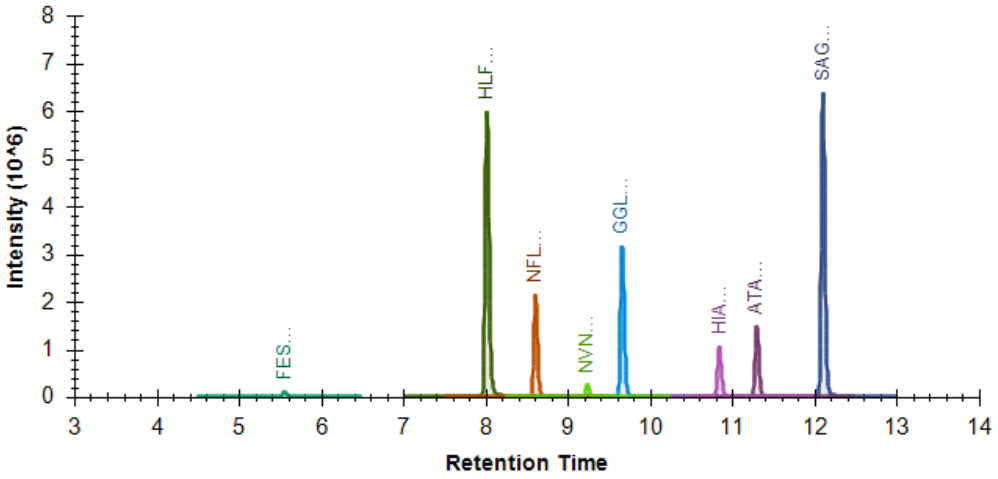



*Figure 51 – UHPLC-HRMS analysis of purified concatemer "Construct 2-3" to confirm concatemer sequence. Purified concatemer was diluted to 0.1 mg/mL with 50 mM TEAB; pH 9.2 and digested with trypsin (trypsin:protein ratio 1:20) for 1 h at 37 °C under 300 rpm orbital agitation.*

### 3.5. Production, purification and characterization of the $^{15}N$ labelled concatemer

The concatemer encoded by the "Construct 2-3" was successfully produced and purified and its sequence was validated by mass spectrometry. Next, the developed strategy was applied to produce a $^{15}N$ isotope-labelled version of this concatemer. To do so, LB and TB used as bacteria growth culture media were replaced by Bioexpress Cell Growth Media (U-15N, 98%). In this culture medium, at least 98 % of the nitrogen corresponds to stable isotope nitrogen-15 ($^{15}N$). Bacteria thus used $^{15}N$ labelled-nutrients to grow and synthetize their own molecules, including the concatemers. With this strategy, a uniform $^{15}N$-labelling can be obtained (Filipp *et al*, 2009).

Pre-culture and culture of transformed bacteria with 'Construct 2-3" were performed with this isotope-labelled medium to an isotopic enrichment as high as possible. A 4 h induction with 1 mM IPTG at 37 °C was performed and the insoluble fraction of obtained cell lysates was purified with Ni-affinity chromatography. Purified isotope-labelled concatemer concentration was then measured based on 280 nm UV light absorbance. Concatemer purity was characterized with SDS-PAGE and gel-image analysis (Figure 52). A total of 2.0 mg of concatemers was successfully produced with a purity of 61.5 % (Figure 52).

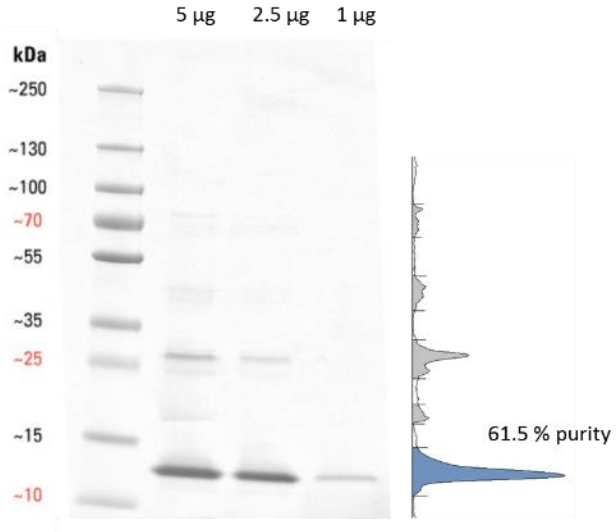



*Figure 52 – SDS-PAGE analysis of purified $^{15}N$ isotope-labelled concatemer encoded by "Construct 2-3". Image analysis of the lane corresponding to 5 µg proteins was performed with ImageJ software to determine the concatemer purity*

The production yield, is lower for the isotope-labelled production of concatemer encoded by "Construct 2-3" when compared to the unlabelled one. With the isotope-labelled growth medium, a total of 2.0 mg of concatemers were produced and purified whereas 3.6 mg were obtained with classical growth medium. These results were somewhat expected as the $^{15}N$ isotope-labelled growth medium is known to be poorer in nutrients than the classical TB growth medium (Berthold *et al*, 2011). The final bacterial biomass obtained at the end of the 4 h induction time in labelled growth medium was less than half of the biomass obtained in classical conditions. However, comparing cell lysates normalized based on cell density, a higher expression yield "per cell" was obtained with labelled growth medium (data not shown).

Purified $^{15}N$ isotope-labelled concatemer "Construct 2-3" was digested with trypsin and analysed by HRMS. The $^{15}N$ isotope labelling introduced a mass shift for the different transitions of each peptide. This mass shift is proportional to the number of nitrogen atoms in the peptide. To illustrate this, an example of this mass shift caused by $^{15}N$ isotope labelling is presented in Figure 53 for the GGLEPINFQTAADQAR peptide. As observed in this figure, the introduced mass shift is sufficient to

eliminate any risk of false positive signal coming from the internal standard. This question is extensively addressed in a research article published in *Food Chemistry* (Gavage *et al*, 2020a) and exposed, in section 4, page 157.

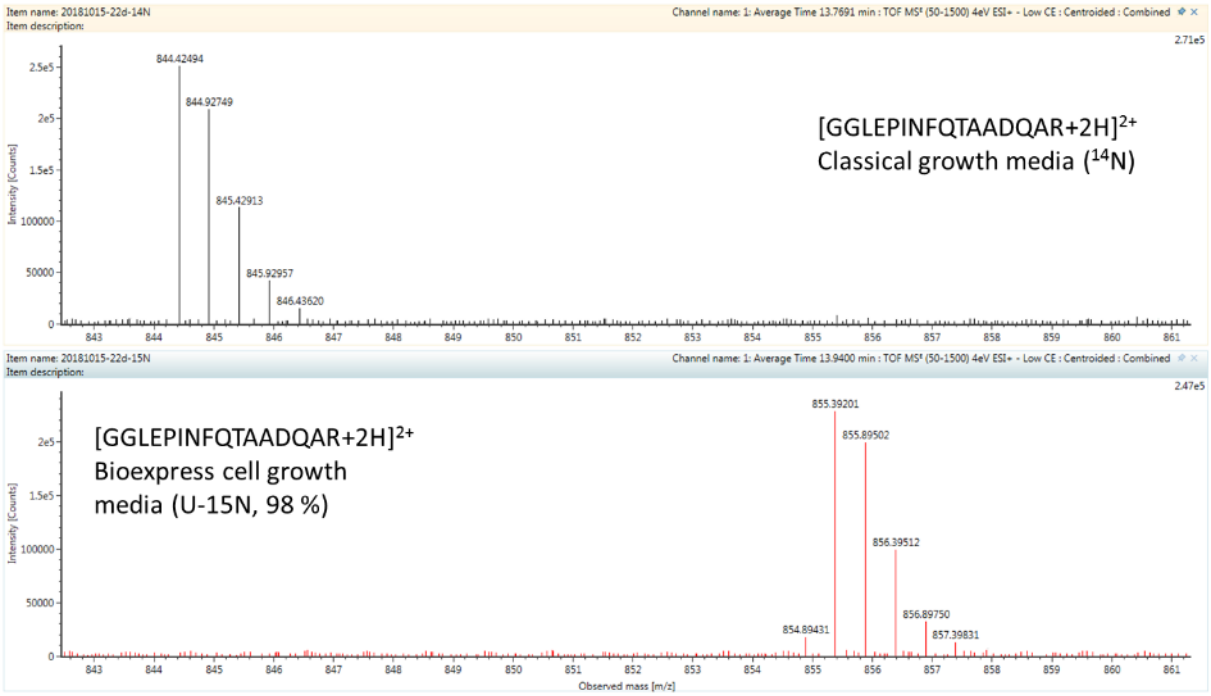



*Figure 53 – Example of mass shift introduced by $^{15}N$ isotope labelling. The spectrum at the top corresponds to peptide GGLEPINFQTAADQAR obtained from the trypsin digestion of concatemer Construct 2-3 produced in classical growth medium whereas the spectrum at the bottom corresponds to concatemer produced in $^{15}N$ isotope labelled growth medium.*

The sequence of the $^{15}N$ isotope-labelled concatemer was verified by HRMS. The 8 egg peptides biomarkers were identified, considering the introduced mass shift (Figure 54).

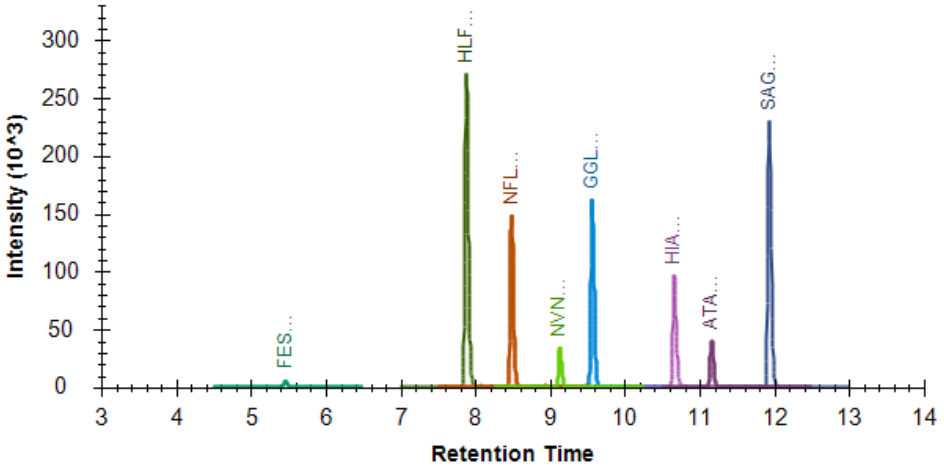



*Figure 54 - UHPLC-HRMS analysis of purified $^{15}N$ isotope-labelled concatemer "Construct 2-3" to confirm the concatemer sequence. Purified concatemer was diluted to 0.1 mg/mL with 50 mM TEAB pH 9.2 and digested with trypsin (trypsin:protein ratio 1:20) for 1 h at 37°C under 300 rpm orbital agitation.*

HRMS was also used to evaluate the isotopic enrichment of the $^{15}N$ isotope-labelled concatemer. The procedure and methods used to evaluate the isotopic enrichment is extensively developed in the research article published in *Food Chemistry* (Gavage *et al*, 2020a) and exposed in section 4. Briefly,

the proportion of the fully $^{15}N$ isotope-labelled version of each tryptic peptide was estimated by comparing the intensities of the monoisotopic peak and those of its isotope containing only one $^{14}N$. For the GGLEPINFQTAADQAR example presented in Figure 53, these peaks correspond to m/z = 855.39201 and m/z = 854.89431, respectively. The relation between the labelling proportion and the number of nitrogen atoms in the peptides follows an exponential decay. The associated exponential decay constant corresponds to the natural logarithm of the isotopic enrichment. Indeed, for a given isotopic enrichment ($\varphi$), the proportion of fully $^{15}N$ labelled peptide with $n$ nitrogens is given by $\varphi^n$, which can be transformed into $e^{\ln(\varphi)*n}$. Isotopic enrichment is deduced from this mathematical transformation by equating $\ln(\varphi)$ to experimentally obtained exponential arguments (-0.00449). This result gave an isotopic enrichment of 99.6 %, a value in agreement with the > 98 % isotopic enrichment of the growth medium.

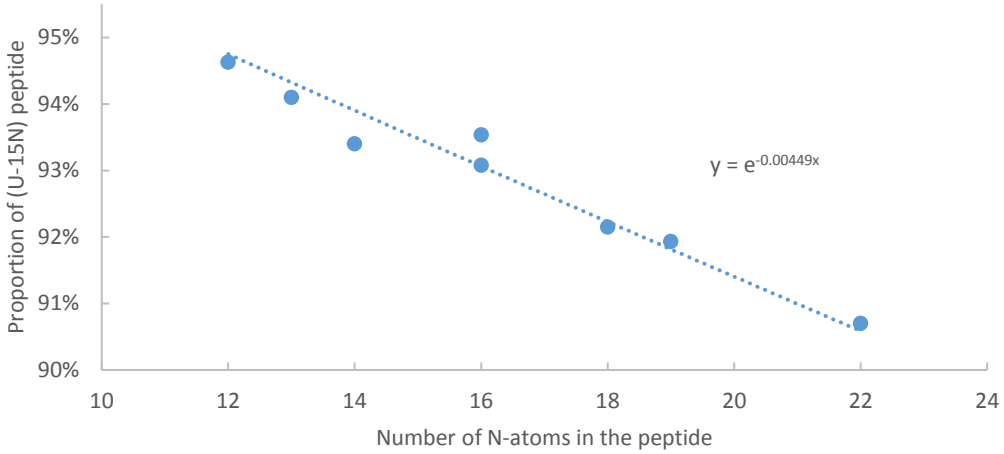



*Figure 55 - Evaluation of the isotopic enrichment for the purified $^{15}N$ isotope-labelled concatemer "Construct 2-3" based on the proportion of fully $^{15}N$ labelled of each identified tryptic peptide*

Altogether, the optimisation of the protocol developed for the protein expression and the isotope labelling strategies resulted in the successful production and purification of a first $^{15}N$ isotope-labelled concatemer made of 8 egg peptide biomarkers.

As the final objective was to produce a single concatemer containing peptide biomarkers of the four targeted allergenic ingredients (egg, milk, peanut and hazelnut). Kaatje Van Vlierberghe during her PhD thesis (PhD ongoing at ILVO under the "Allersens" project), developed the UHPLC-MS/MS method and refined the list of 55 potential peptide biomarkers identified by HRMS analysis of several food processed test matrices to a final list of 19 peptide biomarkers listed in Table 7. These 19 peptide biomarkers were thus considered for the production of a $^{15}N$ isotope-labelled concatemer that could be used as an internal standard for the simultaneous quantification of the four allergenic ingredients.

### 3.6. Application of the protocol for final concatemer production

Gathered expertise in concatemer development and the final list of 19 peptide biomarkers were combined for the design and production of a concatemer containing these 19 peptides and that could be used as an internal standard for the simultaneous quantification of the four allergenic ingredients (peanut, milk, egg and hazelnut). To maximize the chances to obtain a concatemer with a high level of expression, four constructs were designed in parallel, all of them containing the 19 peptide biomarkers (Table 7).

In “Construct 2-1”, an undesired translation initiation site was accidentally introduced. This site was located at the N-term side of the sequence coding for the GFP included in that construct. It was observed that the protein corresponding to the use of this initiation site had a higher expression level than the complete concatemer. It was therefore decided to use the sequence of this GFP translation initiation site in two out of the four new constructs. Hence, the initiation sites of “Construct 3-1” and “Construct 3-2” were corresponding to the previously used GST N-term sequence whereas the initiation sites “Construct 3-3” and “Construct 3-4” were corresponding to the GFP N-term sequence.

The four constructs were respecting the same design criteria. They were all containing the 19 peptide biomarkers (in different order but with an alternation between hydrophilic and hydrophobic peptides), DNA sequences were optimized to avoid strong interactions between nucleotides affecting the mRNA secondary structure and DNA sequences were analysed to identify and eliminate the presence of any potential undesired translation initiation site.

The preliminary assays with a 100 ml batch mode culture with a 4 h induction with IPTG at 37°C were performed for the four constructs. The effect of *E. coli* strains (BL21 (DE3) and BL21 STAR (DE3)) was first investigated (Figure 56).

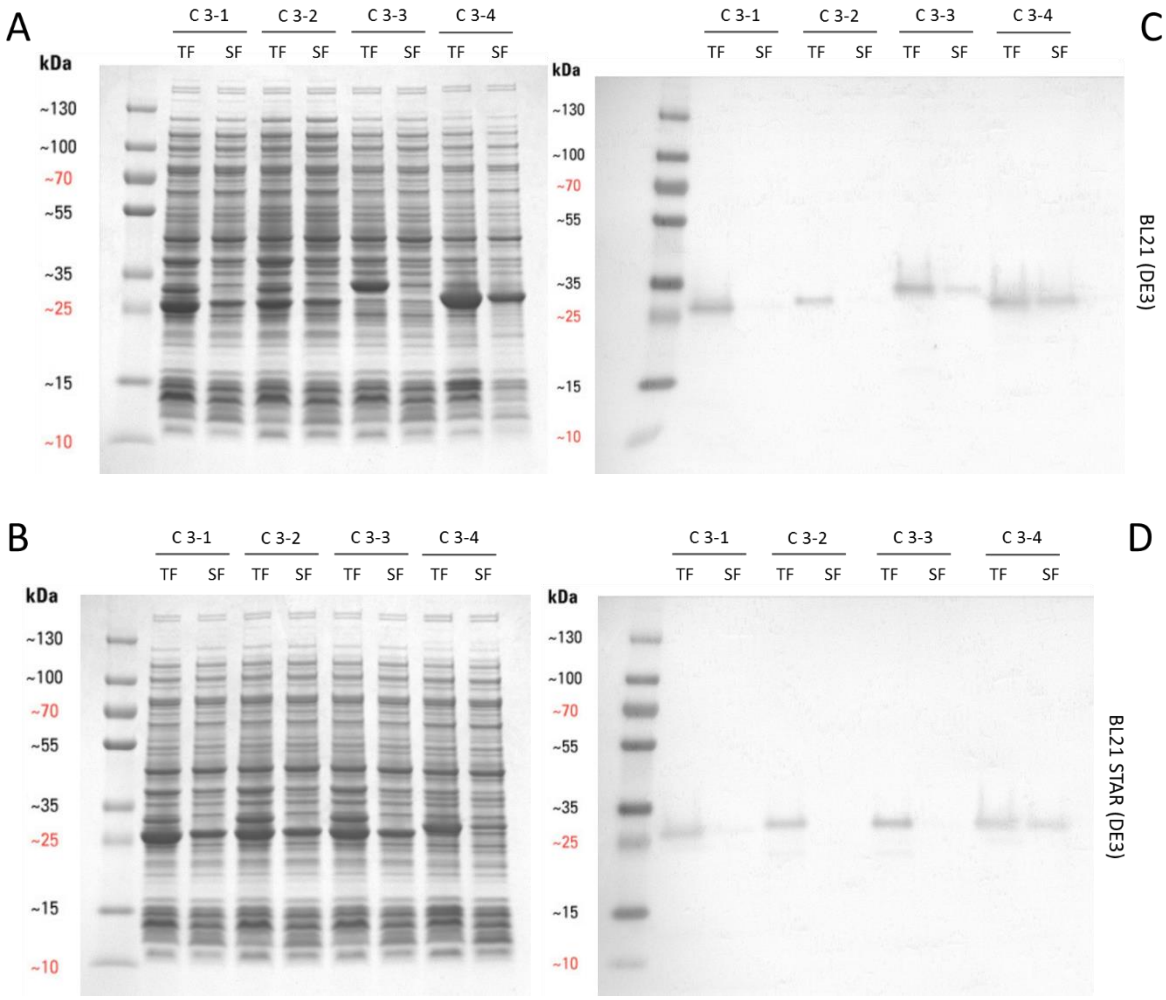



*Figure 56 – Effect of the bacterial strains (BL21 (DE3) (A and C) and BL21 STAR (DE3) (B and D)) and of the different constructs (C3-1 to C3-4) on the expression of the concatemer. The bacteria were grown in a 100 ml batch mode culture at 37°C and the expression of the chimeric protein was induced in the presence of 1 mM IPTG for 4 h at the same temperature. Harvested cells were analysed by (A and B) SDS-PAGE with Coomassie blue staining and (C and D) Western Blot using an anti-Poly-H IgG conjugated to alkaline phosphatase.*

These results show that the different concatemers were correctly expressed. A higher expression yield was also observed for the BL21 (DE3) *E. coli* strain transformed with "Construct 3-4". Moreover, a significant part of the concatemer was found in a soluble form. This solubility was attractive because it simplified the subsequent purification procedure, avoided the consumption of relatively expensive GuHCl, facilitating the use of the protein for future applications.

An unexplained mass shift was observed for "Construct 3-3" expressed *E. coli* BL21 (DE3) strain (Figure 56 C). A slight difference could be observed with "Construct 3-1" or "Construct 3-2" since the primer sequence of "Construct 3-3" (corresponding to GFP N-term sequence) is different from these two constructs (corresponding to GST N-term sequence). However, "Construct 3-3" and "Construct 3-4" should have exactly the same molecular weight, since the order of the peptide biomarkers is the only difference between these two constructs. Protein resistance to the denaturation performed before the migration in the SDS gel could have been an explanation but the mass shift is not observed when "Construct 3-3" was expressed *E. coli* BL21 STAR (DE3) strain.

The "Construct 3-4" was therefore selected and different expression conditions were tested to maximize both the protein expression yield and solubility. The antibiotic concentrations (100 or 200 µg/µl) to increase the selection pressure and induction conditions (16 h at 25°C and 4 h at 37 °C) were studied (Figure 57). The 16 h at 25°C induction condition was investigated instead of the previously considered 16 h at 18°C and 4 h at 30°C conditions. From previous experiment, it was indeed observed that the condition 16 h at 18°C had a positive effect on concatemer solubility but a low final biomass level was obtained and that the 4 h at 37°C condition had a limited effect on concatemer solubility and a lower final biomass level compared to the 4 h at 37°C induction condition. The 16 h at 25°C induction condition was therefore considered as a compromise between 16 h at 18°C and 4 h at 30°C conditions to obtain a soluble concatemer and a high biomass level.

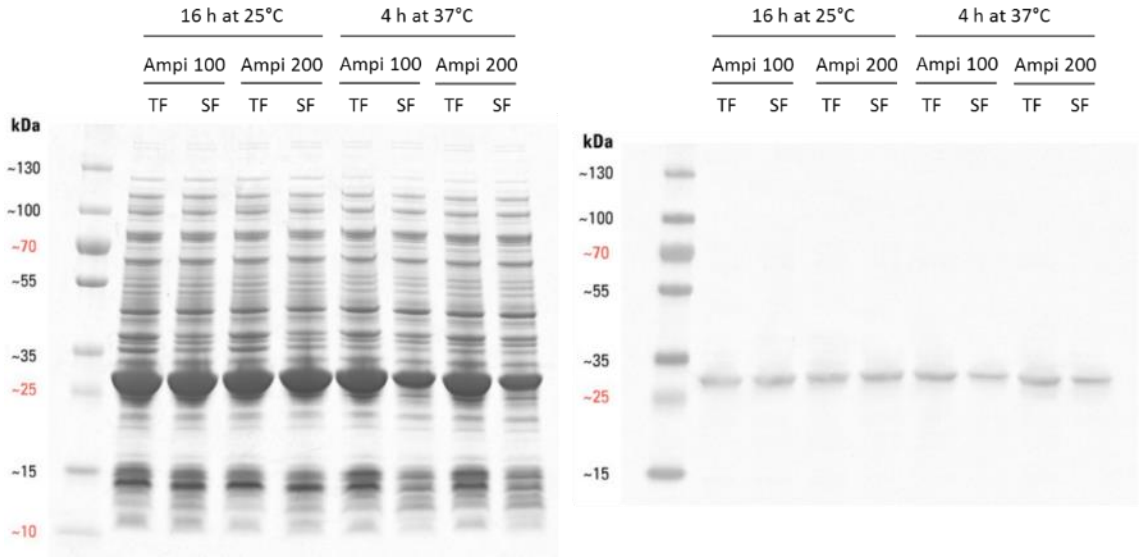



*Figure 57 – Effects of induction conditions (16 h at 25°C and 4 h at 37°C) and ampicillin concentration (100 µg/ml or 200 µg/ml) on the expression of the concatemer encoded by "Construct 3-4". BL21 (DE3) E. coli cells were grown in a 100 ml batch mode culture at 37°C and the expression of the chimeric protein was induced in the presence of 1 mM IPTG. Harvested cells were analysed by SDS-PAGE with Coomassie blue staining and Western Blot using an anti-Poly-H IgG conjugated to alkaline phosphatase.*

We did not see any obvious effect of the antibiotic concentration on the concatemer abundance, while a strong concatemer over-expression with a higher solubility was observed for the 16 h induction at 25°C. These conditions were thus selected for a 1000 ml batch mode production followed by a Ni-affinity chromatographic purification of the soluble fraction of cell lysates.

Elution fractions containing the concatemer, identified based on SDS-PAGE analysis, were pooled. Concatemer concentration was assesses by measuring the 280 nm UV light absorbance and concatemer purity was characterized by SDS-PAGE and gel-image analysis. A total of 21.6 mg of concatemer with a 92 % purity was successfully produced and purified. The purified protein quantity could have been even higher since the 1 ml purification chromatography column was saturated. Indeed, a significant amount of concatemer was observed in purification flow-through. However, the purified concatemer quantity was in agreement with the resin capacity provided by the supplier specifications (20 mg of protein per resin ml).

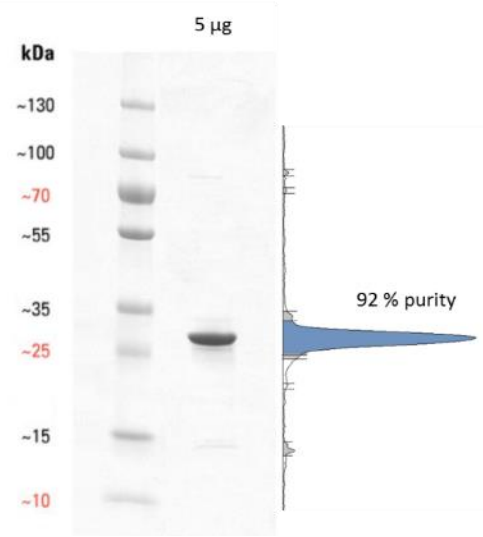



*Figure 58 - SDS-PAGE analysis of purified concatemer encoded by "Construct 3-4". Image analysis of a lane corresponding to 5 µg of proteins was performed with ImageJ software to determine the concatemer purity*

UHPLC-HRMS analysis was used to validate the protein sequence. The purified concatemer was digested with trypsin and the obtained peptides were analysed. As presented in Figure 59, all 19 peptide biomarkers were identified.

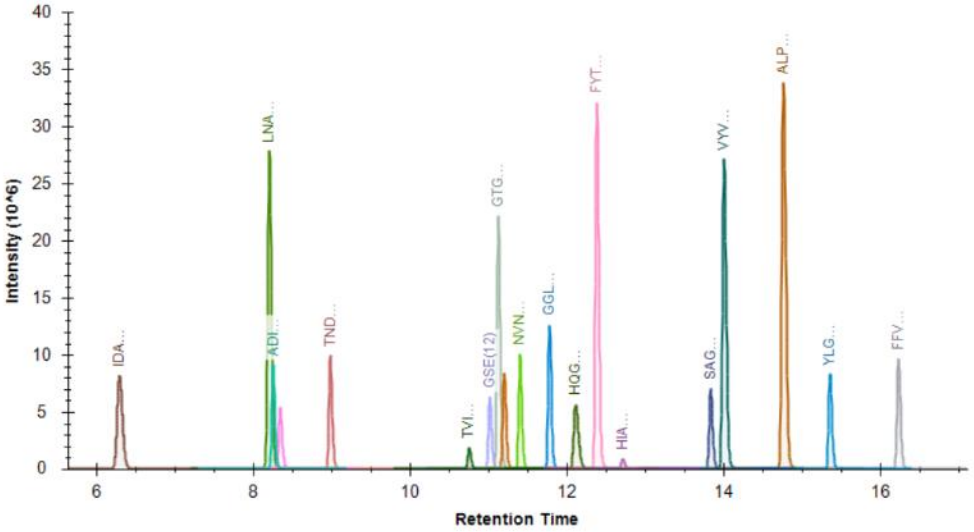



*Figure 59 - UHPLC-HRMS analysis of purified concatemer "Construct 3-4" to confirm concatemer sequence. Purified concatemer was diluted to 0.1 mg/mL with 50 mM TEAB pH 9.2 and digested with trypsin (trypsin:protein ratio 1:20) for 1 h at 37°C under 300 rpm orbital agitation.*

In conclusion, the different developments exposed and described in this chapter led to produce a concatemer with a high expression yield and containing the 19 peptide biomarkers allowing the simultaneous analysis of the 4 allergenic ingredients considered in this project (egg, milk, peanut and hazelnut). In addition, let's mention that one mg of concatemer is sufficient to perform more than a hundred of analyses, making this approach cost-effective when compared to synthetic peptides.

As demonstrated with "Construct 2-3", the isotope labelling strategy based on the use of $^{15}N$ isotope-labelled cell growth medium (Bioexpress Cell Growth Media (U-15N, 98%)) enabled the production of concatemers with a high and sufficient isotopic enrichment. The mass shift introduced by the $^{15}N$ isotope-labelling prevents any risk of false positive signal.

The ultimate step was the implementation of the developed strategy for $^{15}N$ isotope-labelled concatemer "Construct 3-4" production and purification.

### 4. Comparative study of concatemer efficiency as an isotope-labelled internal standard for allergen quantification (Gavage *et al*, 2020a)

A research manuscript was dedicated to this $^{15}N$ isotope-labelled concatemer. The manuscript was published in *Food Chemistry* Journal. Production, purification and characterization of the concatemer are detailed in this article. The use of this technology as an isotopically labelled-internal standard for the quantification of food allergens in processed food matrices was also evaluated. Moreover, performance of the concatemer was compared to the use of labelled synthetic peptides and a $^{15}N$ isotope-labelled protein (β-lactoglobulin, a milk protein that has been recombinantly produced) in the analysis of three food matrices spiked with food allergen extracts.

Food Chemistry 332 (2020) 127413

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

# Comparative study of concatemer efficiency as an isotope-labelled internal standard for allergen quantification

Maxime Gavage[a,b], Kaatje Van Vlierberghe[c], Christof Van Poucke[c], Marc De Loose[c], Kris Gevaert[d,e], Marc Dieu[b,f], Patsy Renard[b,f], Thierry Arnould[b], Patrice Filee[a], Nathalie Gillard[a,*]

[a] CER Groupe, Rue du Point du Jour 8, 6900 Marloie, Belgium
[b] Laboratory of Biochemistry and Cell Biology (URBC), Namur Research Institute for Life Sciences (NARILIS), University of Namur, 61, Rue de Bruxelles, 5000 Namur, Belgium
[c] Flanders Research Institute for Agriculture, Fisheries and Food (ILVO), Brusselsesteenweg 370, 9090 Melle, Belgium
[d] VIB Center for Medical Biotechnology, Albert Baertsoenkaai 3, B9000 Ghent, Belgium
[e] Department of Biomolecular Medicine, Ghent University, Albert Baertsoenkaai 3, B9000 Ghent, Belgium
[f] MaSUN, Mass Spectrometry Facility, University of Namur, 61, Rue de Bruxelles, 5000 Namur, Belgium

ARTICLE INFO

*Keywords:*
Food allergen analysis
Mass spectrometry
Isotope dilution
Isotope-labelled internal standard
Isotope-labelled concatemer

ABSTRACT

Mass spectrometry-based methods coupled with stable isotope dilution have become effective and widely used methods for the detection and quantification of food allergens. Current methods target signature peptides resulting from proteolytic digestion of proteins of the allergenic ingredient. The choice of appropriate stable isotope-labelled internal standard is crucial, given the diversity of encountered food matrices which can affect sample preparation and analysis. We propose the use of concatemer, an artificial and stable isotope-labelled protein composed of several concatenated signature peptides as internal standard. With a comparative analysis of three matrices contaminated with four allergens (egg, milk, peanut, and hazelnut), the concatemer approach was found to offer advantages associated with the use of labelled proteins, ideal but unaffordable, and circumvent certain limitations of traditionally used synthetic peptides as internal standards. Although used in the proteomic field for more than a decade, concatemer strategy has not yet been applied for food analysis.

## 1. Introduction

Food allergy is defined as an adverse health effect arising from a specific reproducible immune response that occurs on exposure to a given food (Boyce et al., 2011). Several studies indicate an increase in the prevalence of food allergy with nearly 5% of adults and 8% of children being affected (Sicherer & Sampson, 2014). Given the absence of accepted treatment, the current solution for allergic patients relies on allergen avoidance to circumvent allergic reactions. However, this essentially requires correct food labelling and efficient risk management from food business operators to reduce the risk of contamination by allergens to acceptable levels. European legislation (Regulation [EU] No 1169/2011) requires the labelling of 14 allergenic ingredients when they are part of a foodstuff recipe. However, this legislation does not cover the presence of hidden allergens that are due to cross-contamination during food processing. Even if strongly requested by food producers and control laboratories, no harmonized regulatory framework for managing hidden allergens or action thresholds have been enacted in Europe. Some countries have set legal thresholds but with a high disparity among allergens and among countries (Planque et al., 2019). A quantitative risk assessment was also developed by VITAL® (Voluntary Incidental Trace Allergen Labelling) combining reference doses and exposure (Allen et al., 2014). The thresholds for allergenic proteins in food are based on clinical data and are indicators of the action levels, expressed as the total protein amount of the allergenic food (mg), below which only the most sensitive allergic subjects might react (1% of allergic patients or 5% of them for the less common foods). These values are often used by laboratories as a targeted limit of quantification (LOQ) in the absence of legal thresholds.

The development of a quantitative allergen risk assessment requires quantitative allergen analysis. During the last decade, mass spectrometry became the method of choice for allergen analysis (Ahsan, Rao, Gruppuso, Ramratnam, & Salomon, 2016). Allergen analysis by mass spectrometry is predominantly performed by specific analysis of peptides obtained by an enzymatic digestion of the proteins of the sample, including the proteins of the allergenic ingredients. One of the

* Corresponding author.
*E-mail address:* n.gillard@cergroupe.be (N. Gillard).

https://doi.org/10.1016/j.foodchem.2020.127413
Received 11 March 2020; Received in revised form 15 June 2020; Accepted 22 June 2020
Available online 27 June 2020

advantages of mass spectrometry-based methods is the possibility to simultaneously detect multiple peptides from multiple allergens, thus enabling time- and money-saving multiplexed analysis. Such a targeted approach, named multiple reaction monitoring (MRM), offers high sensitivity and specificity. Targeted proteomics is often used for absolute peptide quantification in combination with isotope dilution, a technique based on the use of an internal standard corresponding to the stable isotope-labelled version of the analyte (Monaci, Losito, De Angelis, Pilolli, & Visconti, 2013; Nitride et al., 2019; Planque et al., 2019). The introduction of this isotope-labelled internal standard corrects for variability and various matrix effects during the actual analysis. Notably, ion suppression effects and, depending on the type of internal standard, matrix effects and analyte loss during sample preparation may be corrected by the use of isotope-labelled internal standards.

Peptides specific for allergen proteins are the analytes in mass spectrometry analysis of food allergens; however, the initial analytes are proteins. Stable isotope-labelled internal standards can therefore adopt different forms. In theory, a stable isotope-labelled protein is the ideal internal standard as, when added to the food that needs to be analyzed, it can correct for sample losses during all the steps of the sample preparation procedure (including protein extraction and digestion), as well as for matrix effects during mass spectrometry analysis. Such an approach was proposed by Newsome and Scholl (Newsome & Scholl, 2013) for the quantification of bovine milk $\alpha_{S1}$-casein in baked goods. The main limitation of this approach, besides technical issues for protein production, is its cost. When one aims at multiplexed analysis, this necessitates the use of multiple isotope-labelled proteins, which is unrealistic for laboratories performing routine analyses (Planque, Arnould, & Gillard, 2017). Therefore, most laboratories rely on stable isotope-labelled synthetic peptides (Boo, Parker, & Jackson, 2018; Henrottin et al., 2019; Planque et al., 2019). However, in food allergen analysis, the initial analytes are proteins. Peptide internal standard and protein analytes can exhibit different behaviors during the extraction, leading to different extraction yield. Moreover, the peptides do not undergo the enzymatic digestion step which is known to be highly affected by the matrix effects (Korte, Oberleitner, & Brockmeyer, 2019).

Here, we implemented an alternative method based on the synthesis of a concatemer used as a stable isotope-labelled internal standard for allergen quantification. This strategy has been well adopted by proteomics researchers, and the concatemers are known as QconCAT (Pratt et al., 2006), but, as far as we know, these molecules have not yet been explored for food analysis. Concatemers are artificial proteins composed of concatenated, proteotypic peptides originating from different proteins of interest. The peptides themselves are typically first identified following mass spectrometry or are predicted from theoretical peptide sequences. Concatemers are typically recombinantly produced in an environment that allows labelling with stable isotopes (e.g., $^{13}C$ or $^{15}N$). In contrast to synthetic peptides, concatemers need to be proteolytically digested to release their peptides, and thus, this peptide release is also affected by the interference caused by the matrix during the digestion step, in a manner similar to the analyte of interest. Another advantage of concatemers is their potential for multiplexing. A single concatemer can be composed of numerous proteotypic peptides and can therefore be used for multiplexed allergen analysis. The limitation of this approach is fixed by the protein size reachable with recombinant protein expression, which is more than 100 kDa (Chambers, Austen, Fulghum, & Kim, 2004). This approach can be cost-effective when compared with using synthetic peptides for multiplexed analysis. For our study, we developed, produced, and purified a $^{15}N$ isotopically labelled concatemer composed of 19 proteotypic peptides, allowing for the analysis of 4 allergenic ingredients (egg, milk, peanut, and hazelnut). We evaluated the performance of this concatemer by the analysis of three uncontaminated food matrices spiked with increasing and defined concentrations (2.5 ppm to 50 ppm, where ppm corresponded to mg of total allergen protein per kg of matrix) of the selected allergen extracts. In addition, we compared the use of the concatemer with that of five synthetic peptides corresponding to tryptic peptides from the four considered allergens and with β–lactoglobulin, a bovine milk protein that was $^{15}N$ isotopically labelled.

## 2. Material and methods

### 2.1. Reagent and materials

Gene synthesis and cloning were ordered from GeneCust (Boynes, France). Acetic acid, ammonium bicarbonate, ampicillin sodium salt, chloramphenicol, dimethyl sulfoxide (DMSO), DL-dithiothreitol (DTT), expression plasmid pET17b(+) Novagen, HiLoad® 26/600 Superdex® 200 pg, imidazole hydrochloride, iodoacetamide (IAA), kanamycin monosulfate, Lennox broth (LB), Ni Sepharose® 6 fast flow GE Healthcare, Origami™ B(DE3) pLysS competent cells Novagen, phenylmethanesulfonyl fluoride (PMSF), Q Sepharose® Fast Flow, select agar, sodium chloride, sodium phosphate dibasic, sodium phosphate monobasic, tetracycline hydrochloride, tetraethylammonium bicarbonate (TEAB), trypsin from bovine pancreas, tris(hydroxymethyl)aminomethane (Tris) and urea were obtained from Sigma-Aldrich (Bornem, Belgium). One Shot™ BL21(DE3) chemically competent *Escherichia coli*, isopropyl β-D-thiogalactopyranoside (IPTG), SnakeSkin™ dialysis tubing, 3.5 K MWCO, 22 mm were purchased from Thermo Fisher Scientific (Waltham, MA, USA). Bioexpress cell growth media (U-15 N, 98%) (10x concentrate) was obtained from Buchem B.V. (Apeldoorn, the Netherlands), Trypsin Gold, Mass Spectrometry Grade from Promega (Madison, WI, USA), 4–20 Mini-PROTEAN® TGX™ precast protein gels from Bio-Rad (Hercules, CA, USA), Sep-Pak C18 6 cc Vac solid-phase extraction (SPE) cartridges from Waters (Milford, MA, USA), and 0.2 μm acrodisc syringe filters with supor membrane from Pall Corporation (Port Washington, NY, USA). Water, acetonitrile (ACN), and formic acid (FA) were obtained from Biosolve (Valkenswaard, the Netherlands). Labelled synthetic peptides ADIYTEQ**V**[$^{13}C_5{}^{15}N$]GR, FFVAPFPEVFG**K**[$^{13}C_6{}^{15}N_2$], GGLEPIN**F**[*Ring-*$D_5$]QTAADQAR, LS**F**[*Ring-*$D_5$]NPTQLEEQCHI, TANELNLLI**L**[$^{13}C_6{}^{15}N$]R were ordered from Eurogentec (Seraing, Belgium).

Food samples were analyzed by ultra-high performance liquid chromatography–tandem mass-spectrometry (UHPLC MS/MS) using an Acquity liquid chromatograph equipped with a C18 Acquity BEH130 column (2.1 × 150 mm; 1.7 μm) and coupled with a Xevo TQ-S micro triple quadrupole system (Waters, Milford, MA, USA). Characterization of $^{15}N$ isotopically labelled concatemer and β-lactoglobulin was performed by ultra-high performance liquid chromatography–high resolution mass spectrometry (UHPLC-HRMS) using an Acquity liquid chromatograph equipped with a C18 Acquity BEH130 column (2.1 × 150 mm; 1.7 μm) and coupled to a Xevo G2-XS QTof quadrupole time-of-flight system (Waters, Milford, MA, USA).

### 2.2. $^{15}N$ isotopically labelled concatemer production and purification

Design and production of the concatemer were adapted from the method of Pratt (Pratt et al., 2006). The first step focused on concatemer design and the selection of the concatenated peptides. Here, we considered 19 peptides (Table 1) originating from seven proteins of four allergenic ingredients ($\alpha_{S1}$-casein and β-lactoglobulin from cow milk; ovalbumin, ovotransferrin, and vitellogenin-1 from hen's egg; Cor a 9 allergen from hazelnut; and Ara h 1 allergen from peanut). These 19 peptides were selected from a set of relevant peptide biomarkers identified by an empirical approach based on UHPLC-HRMS analysis of incurred and processed samples. The applied food processing steps, sample preparation, and selection criteria have already been detailed in our previous studies (Gavage et al., 2019, 2020; Van Vlierberghe et al., 2020). The peptides were then *in silico* concatenated, and the resulting polypeptide was flanked with an N–terminus initiator sequence including a methionine start and a C–terminus hexahistidine purification

**Table 1**
Targeted peptides and associated multiple reaction monitoring (MRM) parameters for the UHPLC-MS/MS analysis. Isotopically labelled peptides from the (U-$^{15}$N) concatemer are marked with $^{\#}$, those from (U-$^{15}$N) β-lactoglobulin with * and the synthetic peptides with $^{o}$. Product ions selected for internal standard comparison calculation are highlighted in bold.

| Allergen | Protein | Peptide | Precursor (charge state) (*m/z*) | Collision energy (V) | Product ions (fragments) |
|---|---|---|---|---|---|
| Milk | $\alpha_{s1}$-casein | FFVAPFPEVFGK | 693.3 (+ +) | 22 | 992.2 [y9], **921.1 [y8]**, 676.8 [y6] |
| | | (U-$^{15}$N) FFVAPFPEVFGK $^{\#}$ | 699.8 (+ +) | | 1002.1 [y9], **930.0 [y8]**, 683.7 [y6] |
| | | FFVAPFPEVFGK[$^{13}C_6{}^{15}N_2$] $^{o}$ | 697.3 (+ +) | | 1000.1 [y9], **929.0 [y8]**, 684.7 [y6] |
| | | HQGLPQEVLNENLLR | 587.7 (+ + +) | 17 | 872.0 [y7], 790.8 [b7], **758.8 [y6]** |
| | | (U-$^{15}$N) HQGLPQEVLNENLLR $^{\#}$ | 595.6 (+ + +) | | 883.9 [y7], 801.8 [b7], **769.8 [y6]** |
| | | YLGYLEQLLR | 634.7 (+ +) | 18 | **992.2 [y8]**, 771.9 [y6], 658.8 [y5] |
| | | (U-$^{15}$N) YLGYLEQLLR $^{\#}$ | 641.7 (+ +) | | **1004.1 [y8]**, 781.9 [y6], 667.7 [y5] |
| | β-lactoglobulin | IDALNENK | 459 (+ +) | 14 | 803.8 [y7], **688.8 [y6]**, 504.5 [y4] |
| | | (U-$^{15}$N) IDALNENK $^{\#,*}$ | 464.5 (+ +) | | 813.8 [y7], **697.7 [y6]**, 511.5 [y4] |
| | | LSFNPTQLEEQC[+57.1]HI | 858.0 (+ +) | 26 | 1255.4 [y10], 929.0 [y7], **628.2 [y10]** |
| | | (U-$^{15}$N) LSFNPTQLEEQC[+57.1]HI * | 868.4 (+ +) | | 1269.3 [y10], 939.0 [y7], **635.1 [y10]** |
| | | LSF[*Ring-D*$_5$]NPTQLEEQC[+57.1]HI $^{o}$ | 861.5 (+ +) | | 1255.4 [y10], 929.0 [y7], **628.2 [y10]** |
| | | TPEVDDEALEK | 623.7 (+ +) | 20 | 919.0 [y8], 819.8 [y7], **573.1 [y10]** |
| | | (U-$^{15}$N) TPEVDDEALEK $^{\#,*}$ | 629.6 (+ +) | | 927.9 [y8], 827.8 [y7], **578.6 [y10]** |
| | | VYVEELKPTPEGDLEILLQK | 772.2 (+ + +) | 22 | **1026.7 [y18]**, 977.1 [y17], 628.2 [y11] |
| | | (U-$^{15}$N) VYVEELKPTPEGDLEILLQK $^{\#,*}$ | 779.8 (+ + +) | | **1037.1 [y18]**, 987.1 [y17], 634.7 [y11] |
| Egg | Ovalbumin | GGLEPINFQTAADQAR | 844.9 (+ +) | 31 | 1122.2 [y10], 732.8 [y7], **666.7 [y12]** |
| | | (U-$^{15}$N) GGLEPINFQTAADQAR $^{\#}$ | 855.8 (+ +) | | 1138.1 [y10], 743.7 [y7], **675.7 [y12]** |
| | | GGLEPINF[*Ring-D*$_5$]QTAADQAR $^{o}$ | 847.4 (+ +) | | 1127.2 [y10], 732.8 [y7], **669.2 [y12]** |
| | | HIATNAVLFFGR | 673.8 (+ +) | 24 | **1096.3 [y10]**, 1025.2 [y9], 924.1 [y8] |
| | | (U-$^{15}$N) HIATNAVLFFGR $^{\#}$ | 682.7 (+ +) | | **1110.2 [y10]**, 1038.1 [y9], 936.0 [y8] |
| | Ovotransferrin | SAGWNIPIGTLIHR | 512.6 (+ + +) | 15 | **907.1 [y8]**, 696.8 [y6], 538.7 [y4] |
| | | (U-$^{15}$N) SAGWNIPIGTLIHR $^{\#}$ | 519.6 (+ + +) | | **920.0 [y8]**, 707.8 [y6], 547.6 [y4] |
| | | FYTVISSLK | 529.6 (+ +) | 17 | **747.9 [y7]**, 646.8 [y6], 547.7 [y5] |
| | | (U-$^{15}$N) FYTVISSLK $^{\#}$ | 534.6 (+ +) | | **755.9 [y7]**, 653.8 [y6], 553.6 [y5] |
| | Vitellogenin-1 | NVNFDGEILK | 575.1 (+ +) | 17 | **936.0 [y8]**, 674.8 [y6], 559.7 [y5] |
| | | (U-$^{15}$N) NVNFDGEILK $^{\#}$ | 581.6 (+ +) | | **946.0 [y8]**, 681.7 [y6], 565.6 [y5] |
| | | TVIVEAPIHGLK | 639.3 (+ +) | 20 | 865.0 [y8], 735.9 [y7], **664.8 [y6]** |
| | | (U-$^{15}$N) TVIVEAPIHGLK $^{\#}$ | 646.7 (+ +) | | 875.9 [y8], 745.8 [y7], **673.8 [y6]** |
| Hazelnut | Allergen Cor a 9 | ADIYTEQVGR | 576.6 (+ +) | 17 | 852.9 [y7], **689.7 [y6]**, 588.6 [y5] |
| | | (U-$^{15}$N) ADIYTEQVGR $^{\#}$ | 583.6 (+ +) | | 863.8 [y7], **699.7 [y6]**, 597.6 [y5] |
| | | ADIYTEQV[$^{13}C_5{}^{15}N$]GR $^{o}$ | 579.6 (+ +) | | 858.9 [y7], **695.7 [y6]**, 594.6 [y5] |
| | | ALPDDVLANAFQISR | 815.9 (+ +) | 24 | 907.0 [y8], 835.9 [y7], **723.8 [y13]** |
| | | (U-$^{15}$N) ALPDDVLANAFQISR $^{\#}$ | 825.8 (+ +) | | 919.9 [y8], 847.9 [y7], **732.7 [y13]** |
| | | LNALEPTNR | 514.6 (+ +) | 15 | 729.8 [y6], 616.6 [y5], **487.5 [y4]** |
| | | (U-$^{15}$N) LNALEPTNR $^{\#}$ | 521.5 (+ +) | | 739.7 [y6], 625.6 [y5], **495.5 [y4]** |
| | | TNDNAQISPLAGR | 679.2 (+ +) | 21 | 713.8 [y7], **600.7 [y6]**, 513.6 [y5] |
| | | (U-$^{15}$N) TNDNAQISPLAGR $^{\#}$ | 688.7 (+ +) | | 723.8 [y7], **609.6 [y6]**, 521.6 [y5] |
| Peanut | Allergen Ara h 1 | GSEEEDITNPINLR | 794.3 (+ +) | 26 | 828.0 [y7], 726.8 [y6], **612.7 [y5]** |
| | | (U-$^{15}$N) GSEEEDITNPINLR $^{\#}$ | 803.8 (+ +) | | 839.9 [y7], 737.8 [y6], **621.7 [y5]** |
| | | GSEEEGDITNPINLR | 822.9 (+ +) | 26 | 828.0 [y7], 726.8 [y6], **612.7 [y5]** |
| | | (U-$^{15}$N) GSEEEGDITNPINLR $^{\#}$ | 832.8 (+ +) | | 839.9 [y7], 737.8 [y6], **621.7 [y5]** |
| | | GTGNLELVAVR | 565.2 (+ +) | 18 | 800.0 [y7], 686.8 [y6], **557.7 [y5]** |
| | | (U-$^{15}$N) GTGNLELVAVR $^{\#}$ | 572.6 (+ +) | | 809.9 [y7], 695.8 [y6], **565.7 [y5]** |
| | Allergen Ara h 3 | TANELNLLILR | 635.8 (+ +) | 21 | 984.2 [y8], 855.1 [y7], **741.9 [y6]** |
| | | TANELNLLIL[$^{13}C_6{}^{15}N$]R $^{o}$ | 639.2 (+ +) | | 991.2 [y8], 862.1 [y7], **748.9 [y6]** |

tag (His-tag). Hydrophobicity of each of the 19 peptides was evaluated based on their grand average of hydropathy (GRAVY) parameter. Hydrophobic and hydrophilic peptides were alternated in the concatemer sequence to avoid the formation of high hydrophobic clusters that can interfere with solvent accessibility of concatenated peptides and thus with their subsequent proteolysis during the sample preparation. Translation-associated aspects such as tRNA-mediated codon usage bias and mRNA secondary structure, known to impact the translation process (Gorochowski, Ignatova, Bovenberg, & Roubos, 2015), were also considered. Visual Gene Developer (University of California-Davis, Davis, CA, USA) was used to predict and optimize the mRNA secondary structure. The *in silico* designed DNA construct was finally chemically synthesized and cloned into the pET17b(+) expression vector using *NdeI* and *XhoI* restriction sites to give the pET17b(+)-concat1.

The *E. coli* BL21(DE3)/pET17b(+)-concat1 strain was inoculated in a 30 mL starter culture of $^{15}$N labelled media (Bioexpress cell growth media [U-15 N, 98%] with 100 µg/mL ampicillin) and grown overnight at 37 °C under 300 rpm orbital shaking. Cells were harvested by centrifugation (4000 × g, 5 min) and the pellet was resuspended in 1 mL of $^{15}$N labelled media. Next, a volume of 660 µl of this bacterial suspension was used to inoculate a 1L $^{15}$N labelled main culture. This culture was grown at 37 °C under 300 rpm orbital shaking until the optical density at 600 nm reached 0.6–0.8. Concatemer expression was next induced with 1 mM IPTG and cells were cultured overnight at 25 °C under 300 rpm orbital shaking. Cells were harvested by centrifugation (5000 × g, 15 min) and stored at −80 °C until concatemer purification.

The cell pellet of the 1 L culture was resuspended in 40 mL of lysis buffer (50 mM Tris − 10 mM imidazole - pH 8) with 1 mM PMSF. Cells were disrupted using a Vibra-Cell™ (Sonics, Newtown, CN, USA) ultrasonic probe. The cell lysate was centrifuged twice (40000 × g, 20 min) and filtered through 0.2 µm syringe filters before to be submitted to metal affinity chromatography purification. The protein solution was loaded on a 8 mL Ni Sepharose 6 Fast Flow column equilibrated with lysis buffer. An intermediate washing step was performed in the presence of 20 mM imidazole and the His-tag labelled concatemer was finally eluted by using a linear imidazole gradient from 20 mM to 250 mM. The elution fractions were analyzed on SDS-PAGE (Supplementary data 1). The positive fractions were pooled and dialyzed against the storage buffer (50 mM Tris - pH 8) to eliminate

imidazole.

Total protein concentration was measured by absorbance at 280 nm. A SDS-PAGE/densitometry method based on ImageJ software was used to estimate concatemer purity. A total of 84.5 mg of $^{15}$N isotopically labelled concatemer were produced and purified with an estimated purity higher than 90%. Protein sequences, concentration calculations, and purity estimation are detailed in Supplementary data 3.

### 2.3. $^{15}$N isotopically labelled β-lactoglobulin production and purification

The production of β-lactoglobulin, a cow milk protein, was adapted from the work of Loch and collaborators (Loch et al., 2016) who implemented a method leading to the cytoplasmic accumulation of correctly folded disulfide bond-dependent proteins. Briefly, two mutations (L2A/I3S) were introduced in the β-lactoglobulin to facilitate *in vivo* cleavage of the *N*-terminal methionine allowing for correct protein folding.) Further, the *E. coli* Origami B (DE3) pLysS strain, a glutathione reductase (gor) and thioredoxin reductase (trxB) mutated strain, was used for conducting the cytoplasmic co-expression of the protein of interest with DsbC, an *E. coli* cytoplasmic disulfide bond isomerase. The co-expression was achieved with the same expression vector (pET17b (+)-DsbC-BLg) in which the two genes were transcribed from individual T7 IPTG-inducible promoters.

To achieve the production of $^{15}$N labelled β-lactoglobulin, expression (starter culture, main culture, and IPTG induction) conditions were similar as used for concatemer production. The antibiotics that were used were tailored to 200 μg/mL ampicillin, 34 μg/mL chloramphenicol, 15 μg/mL kanamycin, and 12.5 μg/mL tetracycline, and the IPTG concentration was 0.5 mM. Harvested cells were resuspended in 50 mM phosphate buffer, pH 6.5, with 1 mM PMSF and prepared for protein purification using the same procedure as for the concatemer. The purification of 15 N labelled β-lactoglobulin was performed according to the procedure described by Loch and collaborator (Loch et al., 2016). Briefly, this protocol combines anion-exchange chromatography (Q Sepharose® Fast Flow) with a NaCl linear elution gradient (up to 2 M) followed by size-exclusion chromatography (HiLoad® 26/600 Superdex® 200 pg) in initial conditions (50 mM phosphate buffer, pH 6.5). Eluates of these two purification steps were collected in 1 mL fractions and analyzed on SDS-PAGE (Supplementary data 2).

Total protein concentration was measured by absorbance at 280 nm. A SDS-PAGE/densitometry method based on ImageJ software was used to estimate protein purity. Using this approach, a total of 2.4 mg of $^{15}$N isotopically labelled β–lactoglobulin were produced and purified with an estimated purity higher than 70%. Protein sequences, concentration calculations, and purity estimation are detailed in Supplementary data 3.

### 2.4. Characterization of produced $^{15}$N isotopically labelled proteins

Protein $^{15}$N stable isotope enrichment was evaluated by UHPLC-HRMS analysis of its constitutive tryptic peptides. In separated containers, concatemer and β-lactoglobulin were diluted to 0.1 mg/mL with 50 mM TEAB, pH 9.2, to a final volume of 20 μl. Disulfide bridges of β-lactoglobulin were successively reduced and alkylated with DTT (10 mM final concentration, 45 min incubation at 37 °C under 300 rpm orbital agitation) and IAA (40 mM final concentration, 45 min incubation in the dark at 37 °C under 300 rpm orbital agitation). Concatemer and β-lactoglobulin were then proteolytically digested by adding 0.1 μg of trypsin gold (protein:trypsin ratio of 1:20). Digestion was conducted for 1 h at 37 °C under 300 rpm orbital agitation and stopped by the addition of 1% (final concentration) of FA followed by centrifugation (20000 × g, 5 min). Samples were ten–fold diluted with 5% ACN before UHPLC-HRMS analysis. Peptides (5 μl of sample was injected) were first separated by reverse-phase liquid chromatography using a 20 min water/ACN + 0.1% FA linear gradient from 5% to 40% of ACN. Data was acquired in MS$^E$ mode with 0.3 s scan time within the 50 to 2000 *m/z* mass range. The data were processed using UNIFI software (Waters, Milford, MA, USA) and peptide mapping analysis type with traditional tryptic cleavage rules and setting cysteine carbamidomethylation and $^{15}$N isotope labelling as a fixed modifications.

For each identified tryptic peptide, the most intense charge state was considered to define the $^{15}$N stable isotopic enrichment. The isotopic enrichment or isotope incorporation rate was evaluated for each peptide by comparing the intensity (in counts) of the peak corresponding to the fully $^{15}$N labelled (U–$^{15}$N) peptide with other peaks corresponding to partially $^{15}$N labelled peptides. For practicality, we considered a $^{13}$C natural abundance of 1.1% and neglected hydrogen and oxygen isotopic distributions in our calculations. Furthermore, only peaks corresponding to peptide with 1 $^{14}$N isotope were considered in our calculation. The proportion of U-$^{15}$N peptide was then obtained after comparing the intensity of the peak corresponding to the (U-$^{15}$N & U–$^{12}$C) peptide with the peak corresponding to the [(U-1)-$^{15}$N & U-$^{12}$C] peptide. Protein isotopic enrichment was evaluated with the exponential trend given by the proportion of the U-$^{15}$N version of each peptide considering its nitrogen content.

### 2.5. Food matrices preparation

Three blank food matrices – thus, not contaminated with the considered allergenic ingredients – were prepared to assess the variability due to the food sample used in our study. These blank matrices were baked cookies, chocolate, and freeze-dried cookie dough.

Cookie dough was produced in batches of 3 kg by mixing (Kenwood Major Titanium, Stainless Steel Dough Hook, 15 min, max speed) the following ingredients purchased from a local supermarket in the respective weight proportions as follows: wheat flour (Carrefour type 55)/water (Milli-Q)/olive oil (Bertoli Classico)/salt (sodium chloride ACS, ≥ 99%, Thermo Scientific™)/baking powder (Dr. Oetker Baking)/Sugar (Grand Pont Crystal Sugar): 57%/18%/10%/0.2%/0.8%/14%. The dough was subsequently rolled out to a thickness of 0.5 mm, and cookies with a diameter of 8 cm were pressed out of the dough (weight = 25 ± 2 g). Cookies were baked for 25 min with the following program: 1–10 min: 180 °C heat from above and 180 °C heat from below; 11–25 min: 180 °C heat from above and 160 °C heat from below. This was done to ensure that the warming of the baking plate would not result in uneven cookie baking. Cookies were left at ambient temperatures to cool down, and subsequently milled and sieved (Retsch® ZM 200 ultra-centrifugal mill [Retsch GmbH, Haan, Germany] with a 0.75 mm pore size sieve, 14000 rpm). Cookie powder was stored at 4 °C in the dark until further use.

Cookie dough was produced as described above, rolled out to a thickness of 1 cm, stored at –20 °C, and subsequently freeze-dried. Freeze-dried cookie dough was then milled and sieved (Retch® ZM 200 ultra-centrifugal mill with a 0.75 mm pore size sieve, 14000 rpm). The freeze-dried cookie dough powder was stored at 4 °C in the dark until further use.

Chocolate was made by warming chocolate walsenpowder (90%; Callebaut, Belgium) and cacao butter (10%, Callebaut, Belgium) in a water bath at 40 °C (maximum temperature). The mixture was stirred for 15 min, after which 2% ammonium phosphatide (kindly provided by Palsgaard, Julesminde, Denmark) was added. This mixture was again stirred for 15 min and subsequently poured into chocolate molds, resulting in chocolate chips of around 5 g each. The chocolate was left to cool down and solidify at 4 °C for 2 h, and the chocolate chips were packed under vacuum and stored at 4 °C in the dark until further use.

### 2.6. Sample preparation for UHPLC-MS/MS analysis

Two series of samples were prepared and analyzed to be able to cover the three internal standards. Concatemer and β-lactoglobulin were isotopically labelled with the same strategy ($^{15}$N uniform labelling) and share common tryptic peptides, which cannot be

distinguished after enzymatic digestion. Two series of samples were prepared. Labelled peptides and β-lactoglobulin were spiked in the first series (only one shared peptide LSFNPTQLEEQCHI) and labelled concatemer in the second one. For each series, the three blank matrices (baked cookie, chocolate, and lyophilized unbaked cookie dough) were spiked, before extraction, with the appropriate internal standard and with increasing amounts of a standard extract of the four allergens (milk, egg, peanut, and hazelnut). These allergen amounts corresponded to 0, 2.5, 5, 10, 25 and 50 ppm level points expressed in total allergen protein per matrix kg. For each series, each blank matrix and each point of the allergen curve, three biological sample replicates were prepared and analyzed. Stock solutions containing the four allergen standards at 20 mg/mL were prepared using a similar extraction protocol as that used for the samples (extraction, sonication, and centrifugation; see below). These stock solutions were then combined and diluted in appropriate ratios to spike samples at different contamination levels with a 100 µl volume. Combination and dilution were calculated based on theoretical protein content of standards assuming 100% extraction yield. Each internal standard was spiked at the similar molar level (0.25 nmol) with a 100 µl volume. Then,1 mg/mL stock solutions of the five considered labelled peptides (ADIYTEQ**V**[$^{13}C_5{}^{15}N$]GR, FFVAPFPEVFG**K**[$^{13}C_6{}^{15}N_2$], GGLEPIN**F**[*Ring-*$D_5$]QTAADQAR, LS**F**[*Ring-*$D_5$]NPTQLEEQCHI, and TANELNLLI**L**[$^{13}C_6{}^{15}N$]R) were combined and diluted at the appropriate concentration with 0.1% FA. Concatemer and β-lactoglobulin solutions were also diluted to be spiked at 0.25 nmol level with a 100 µl volume. This level, converted in equivalent allergen ppm, ranged from 10 ppm for abundant proteins, such as $\alpha_{S1}$-casein, to more than 300 ppm for less abundant proteins, such as vitellogenin-1. This estimate was based on the natural abundance of each considered protein in the corresponding allergenic ingredient. Allergen standards and internal standard were added to blank matrices before extraction.

Samples were prepared as previously described (Planque et al., 2016). Briefly, protein from 2 g samples was extracted in 50 mL conical tubes with 20 mL of 200 mM Tris, pH 9.2, 2 M urea by shaking at 20 °C for 30 min (Agitelec, J. Toulemonde, Paris, France) prior to ultrasonic treatment at 4 °C for 15 min. After centrifugation (4660 × g, 10 min), 10 mL of supernatant were diluted in digestion buffer (200 mM ammonium bicarbonate, pH 8.2). Protein disulfide bridges were successively reduced and alkylated with 45 min incubation steps at room temperature with the addition of 1 mL of 200 mM DTT and 1 mL of 400 mM IAA (in the dark). Protein was then enzymatically digested with the addition of 1 mL of trypsin solution (trypsin from bovine pancreas, 1 mg/mL in 50 mM acetic acid, pH 2.8) and incubation for 1 h at 37 °C. The digestion reaction was stopped by adding 300 µl of 20% FA to the samples, which were then centrifuged (4660 × g, 5 min). Obtained peptides were then purified and concentrated using C18 SPE cartridges, which were first conditioned with 18 mL of ACN followed by 18 mL of 0.1% FA before loading of 20 mL of the centrifuged sample. The cartridges were washed with 18 mL of 0.1% FA and eluted in 15 mL conical tubes with 6 mL of 80% ACN and 0.1% FA. A volume of 30 µl of DMSO was added to the sample before evaporation (40 °C under nitrogen flow) to avoid dryness. The pellet was finally dissolved in 600 µl of 5% ACN with 0.1% FA and centrifuged twice (4660 × g, 5 min in conical tube and 20 000 × g, 5 min in 1.5 mL microtube, keeping the supernatant) before UHPLC-MS/MS analysis.

### 2.7. UHPLC-MS/MS analysis and data analysis

The peptides were separated by reverse-phase chromatography online connected to a triple quadrupole mass spectrometer. The following 26 min solvent gradient (solvent A, 0.1% FA and solvent B, ACN and 0.1% FA) was applied to the 20 µl injected sample volume: 0–3 min: 92% solvent A; 3–18 min: linear gradient from 92% to 58% solvent A; 18–22.5 min: 15% solvent A; and 22.5–26 min: 92% solvent A, always at constant 0.2 mL/min flow rate. Eluted peptides were ionized using the positive electrospray source and analyzed in MRM mode. The source gas flow was set at 50 L/h and the source voltage at 2.5 kV for the capillary and 30 V for the cone. The source temperature was set at 150 °C and the desolvation temperature at 400 °C with a gas flow at 1200 L/h. Targeted transitions are summarized in Table 1. For each peptide, three transitions were analyzed, as well as the corresponding transitions for the related isotopically labelled internal standard(s) (peptides, concatemer, and β-lactoglobulin). The transitions were selected beforehand using criteria that included the MS signal intensity and the absence of interference for the three considered matrices. The MS/MS acquisition method was generated using the open source Skyline software (MacLean et al., 2010). The most intense transition was used for internal standard comparison calculation and the two others as confirmatory transitions. Internal standards were compared using the peak area ratio (for the most intense transition) between the peptide from the allergenic ingredient and its corresponding isotopically labelled version from the internal standard.

## 3. Results and discussion

### 3.1. Choice of the isotope labelling strategy

Stable isotope internal standard labelling and associated isotopic enrichment are key elements in the design of quantitative mass spectrometry-based methods. The isotopic enrichment and mass shift combination has to be sufficient to avoid any potential risk of false positive introduction. The resolution of quadrupole analyzers is typically around 1 atomic mass unit (Georgiou & Danezis, 2015). Taking into account that most of the peptide ion precursors carry multiple charges and that peptides contain tens of carbons, which lead to widespread isotopic distribution (see Fig. 1), the mass shift introduced by the stable isotopes has to be sufficient to be able to totally distinguish the natural analyte from its internal standard. Considering these aspects, a mass shift of $m/z \geq 3$ is necessary. Furthermore, attention has to be paid to the actual isotope enrichment. Depending on the labelling strategy, an insufficient isotope enrichment may lead to the introduction of the unlabeled form of the internal standard, thus corresponding to the natural analyte itself and contaminating the quantitative analysis.

Several strategies have been developed to produce isotopically labelled proteins, including selective labelling using auxotrophic *E. coli* strains and growth medium supplemented with isotopically labelled amino acids (Mondal, Shet, Prasanna, & Atreya, 2013) or post-translational protein deuteration (Galan et al., 2018). In this study, we decided to use a rich bacterial cell growth medium specifically designed for $^{15}N$ labeling protein using *E. coli* as a host cell for recombinant protein expression. This original medium is an algal hydrolysate that contains the same level of amino acids as LB medium. This strategy allowed for stable and protein sequence independent labelling (as each amino acid contains at least one nitrogen) with a high isotopic enrichment. As one of the peptide biomarkers selection criteria concerned the actual peptide length (peptides should have at least 8 amino acids), $m/z \geq 3$ mass shift precaution is respected for triply charged precursor. Indeed, selected peptide biomarkers are tryptic peptides, with a lysine or an arginine in C-terminal position, holding two and four nitrogen atoms, respectively.

### 3.2. Characterization of $^{15}N$ isotopically labelled proteins

The isotopic enrichment in the concatemer and β-lactoglobulin was evaluated following analysis of their constitutive tryptic peptides by UHPLC-HRMS. The proportion of the fully $^{15}N$ labelled version of each tryptic peptide was estimated by comparing the intensities of the monoisotopic peak (U-$^{15}N$ & U-$^{12}C$) and those of its isotope containing one $^{14}N$ isotope ([U-1]-$^{15}N$ & U-$^{12}C$). As shown in Fig. 1, the intensities of the peaks from peptides with more than one $^{14}N$ isotope were found to be negligible (relative peak intensity < 1% compared to the [U-$^{15}N$

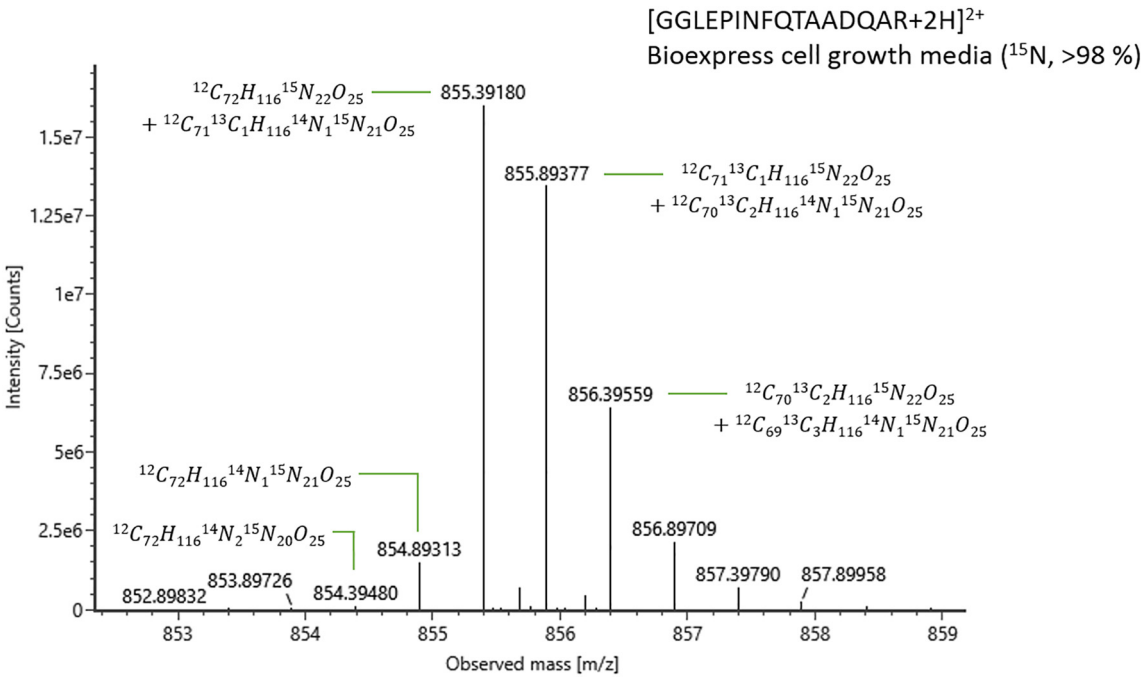



**Fig. 1.** Isotope distribution of the doubly charged ion of the peptide GGLEPINFQTAADQAR (from egg white ovalbumin), one of the 19 peptides of the $^{15}N$ isotopically labelled concatemer. Ignoring oxygen and hydrogen isotopes and peptides with more than two $^{14}N$ isotopes, peaks are annotated with their corresponding peptide formulas.

& U-$^{12}C$] peak). Given the resolution of the MS system (40000), carbon and nitrogen isotopes could not be distinguished. As a result, the monoisotopic peak (U-$^{15}N$ & U-$^{12}C$) was combined with the peak corresponding to the peptide with one $^{14}N$ and one $^{13}C$ isotope ([U-1]-$^{15}N$ & $^{13}C_1$). The proportion of fully $^{15}N$ labelled peptide was evaluated by comparing (U-$^{15}N$ & U-$^{12}C$) and ([U-1]-$^{15}N$ & U-$^{12}C$) peak intensities. The part of the peak intensity corresponding to the (U-$^{15}N$ & U-$^{12}C$) isotope therefore had to first be discriminated from the combined (U-$^{15}N$ & U-$^{12}C$) and ([U-1]-$^{15}N$ & $^{13}C_1$) isotopes' peak intensity. Since isotopes with more than one $^{14}N$ were found to be negligible, we assumed that the ([U-1]-$^{15}N$ & U-$^{12}C$) isotope peak would only correspond to this combination of isotopes. The peak intensity of the ([U-1]-$^{15}N$ & $^{13}C_1$ isotope could therefore be predicted from the ([U-1]-$^{15}N$ & U-$^{12}C$) isotope peak intensity assuming a 1.1% natural abundance of $^{13}C$ isotopes and knowing the number of carbon atoms in the peptide. With this prediction, the (U-$^{15}N$ & U-$^{12}C$) isotope peak intensity could be deduced from the combined isotopes' peak intensity.

The proportion of fully $^{15}N$ labelled peptide was evaluated for all the 19 concatenated tryptic peptides of the concatemer and for all identified tryptic peptides from β-lactoglobulin. As shown in Fig. 2, the relation between the labelling proportion and the number of nitrogen atoms in the peptides follows an exponential decay. The associated exponential decay constant corresponds to the natural logarithm of the isotopic enrichment. Indeed, for a given isotopic enrichment ($\varphi$), the proportion of fully $^{15}N$ labelled peptide with $n$ nitrogens is given by $\varphi^n$, which can be transformed into $e^{\ln(\varphi)*n}$. Isotopic enrichment is deduced from this mathematical transformation by equating $\ln(\varphi)$ to experimentally obtained exponential arguments ($-0.00446$ for the concatemer and –0.00411 for β-lactoglobulin). These results give an isotopic enrichment of 99.5% for the concatemer and 99.6% for β-lactoglobulin, and are in agreement with the greater than 98% isotopic enrichment of the growth medium.

By using a method for efficient isotopic labelling of recombinant protein, we demonstrated that the purified $^{15}N$ isotopically labelled concatemer and β-lactoglobulin internal standards fulfilled the required criteria regarding isotopic enrichment and the introduced mass shift. With this $^{15}N$ uniform labelling strategy, the introduced mass shift was sufficient to distinguish the internal standard from the natural analyte using the quadrupole analyzer. The lowest mass shift corresponded to the double charged FYTVISSLK peptide (from egg white ovotransferrin), one of the 19 concatenated peptides, which contained 10 nitrogen atoms and an associated mass shift of a $m/z$ of 5. Such a mass shift and obtained isotopic enrichment combination prevented the risk of false positive introduction.

### 3.3. Comparison of isotopically labelled internal standards

Performance of the three types of isotopically labelled internal standards (peptides, concatemer, and protein) were evaluated following analysis of three food matrices (baked cookie [cookie], chocolate, and lyophilized unbaked cookie dough [dough]). In theory, a perfect internal standard would have the same exact behavior as its corresponding analyte during sample preparation and analysis. Hence, any analyte loss or matrix effect (during sample preparation or UHPLC-MS/MS analysis) which affects the analyte should equally affect the internal standard. Consequently, for a given natural analyte concentration and internal standard spike level, the signal ratio between a natural analyte and the internal standard would remain constant, independent of analyte losses and matrix effects. The three internal standards considered in this study were compared based on this correlation.

Similar matrix-matched calibration curves were prepared for the three matrices. These curves included a blank and five allergen concentrations ranging from 2.5 to 50 ppm (expressed in mg total allergen protein per kg of matrix), with each sample prepared in triplicate. For each combination of matrix and allergen contamination level, the appropriate internal standard(s) (isotopically labelled peptides and U-$^{15}N$ β-lactoglobulin for the first sample series, and U-$^{15}N$ concatemer for the second one) was spiked at the same concentration. Results are presented separately for each targeted peptide and its corresponding internal standard (five synthetic peptides, 19 allergenic tryptic peptides from U-$^{15}N$ concatemer digestion, and four tryptic peptides from U-$^{15}N$ β-lactoglobulin digestion). Representative peptides of each internal standard are shown in Fig. 3, and complete results are shown in Supplementary data 4. Performance of the different internal standards

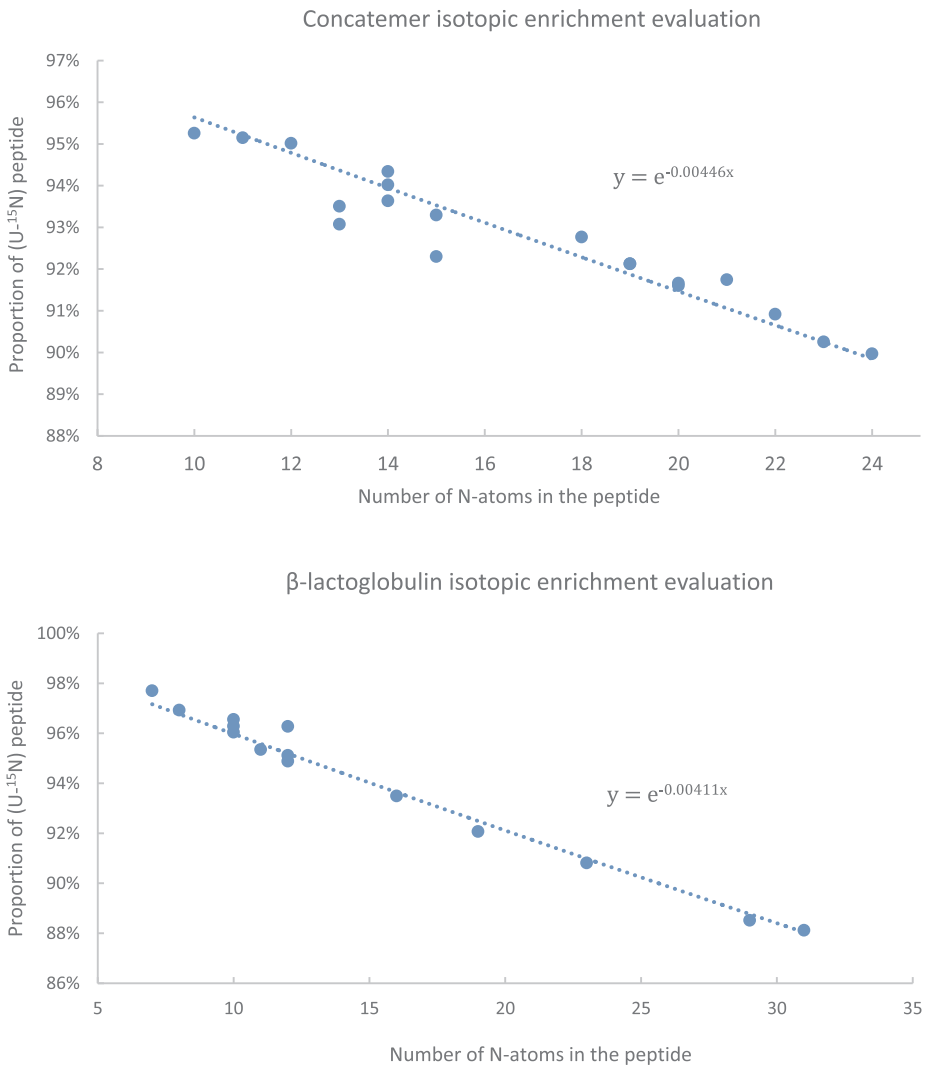



**Fig. 2.** Evaluation of the isotopic enrichment for the purified $^{15}$N isotopically labelled concatemer and β-lactoglobulin. The proportion of fully $^{15}$N labelled peptide was evaluated for each identified tryptic peptide. The relation between the labelling proportion and the number of nitrogen atoms in the peptide follows an exponential decay of which the exponential decay constant corresponds to the natural logarithm of the isotopic enrichment.

were evaluated by comparing the peak area ratio for the most intense transition (highlighted in Table 1) between the analyte and its corresponding internal standard for the three considered matrices. As shown, for a given analyte and internal standard concentration, the signal ratio remained constant when the internal standard was effective. The overlay of the generated linear regression lines was therefore used to evaluate internal standards performance. Overlapping regression lines indicated, for each allergen contamination level, a constant peak area ratio among matrices and thus, an effective internal standard, compensating for matrix effects. In addition to visual evaluation, overlapping regression lines were evaluated using the coefficient of variation (CV) between the slopes of the linear regression lines.

Overall, the best results were obtained for the isotopically labelled protein, U-$^{15}$N β–lactoglobulin. Assuming that recombinant protein folding was similar to the native protein and that the introduced *N*-terminal mutations had no significant impact, as previously demonstrated (Loch et al., 2016), this approach seemed to be the one best suited one for quantifying allergen proteins. Aside from their mass (given the mass shift introduced by isotope labelling), both the analyte protein and the internal standard protein must have had the same properties. This was confirmed by the analysis of four constitutive tryptic peptides from β–lactoglobulin. Regression lines overlapped with all CV values below 15%. This confirmed that the internal standard had efficiently balanced matrix effects during sample preparation and UHPLC-MS/MS analysis, further supported by the fact that the analyte absolute peak area varied by a factor of up to 10 among the three considered matrices, depending on the peptide (data not shown) while the analyte/internal standard peak area ratio remained constant. However, the labelled protein was spiked into the different samples after food processing, which is known to impact peptide detectability and quantification (Korte et al., 2019; Parker et al., 2015). Peptide biomarker selection is therefore a crucial preliminary step in the development of a quantitative method, and selected peptides have to be robust to the food process.

The results obtained with isotopically labelled peptides and the concatemer were less straightforward to interpret. For some targeted peptides, such as LSFNPTQLEEQCHI with labelled peptides, or TNDNAQISPLAGR with the U–$^{15}$N concatemer, the internal standard efficiently compensated for matrix effects with observed CV values below 15%. However, for some other targeted peptides, such as GGLEPINFQTAADQAR with both U–$^{15}$N concatemer and labelled peptides, the analyte and internal standard signal ratio was highly matrix-dependent. In these cases, internal standards did not correctly balance matrix effects, potentially leading to biased allergen quantification. These results are consistent with those reported by Planque and co-workers (Planque et al., 2019). No significant difference was observed for the three

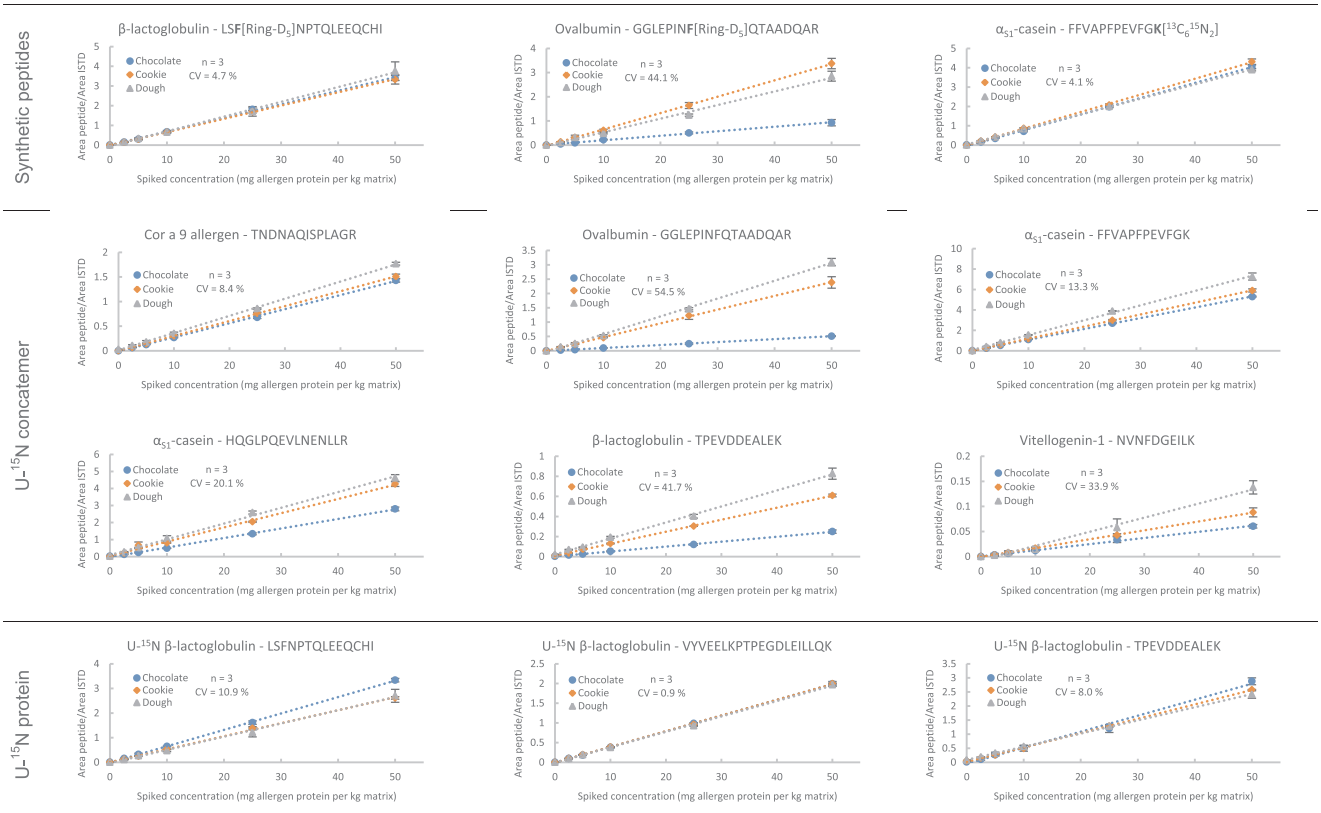



**Fig. 3.** Comparison of three different types of isotopically labelled internal standards for allergen quantification [synthetic peptides, U-$^{15}$N concatemer and U-$^{15}$N protein (β-lactoglobulin)]. For each combination of matrix and allergen contamination level, samples were prepared and analysed in three independent replicates. Results are exposed as the means ± 1 SD of the peak area ratio of the most intense transition (highlighted in Table 1) between the analyte (the allergenic peptide) and the corresponding internal standard. A linear regression line is added for each matrix and the associated coefficient of variation (CV) between the linear regression line slopes is evaluated. For each internal standard, representative peptides are exposed. Full results can be found in Supplementary data 4.

peptides which were common to synthetic peptides and concatemer used as internal standards.

Isotopically labelled peptides are not subject to one of the crucial steps during sample preparation, this being the proteolytic digestion with trypsin. The composition of the food matrix directly impairs the efficiency of enzymatic digestion at least in two different ways. First, different matrices have different protein concentrations, directly affecting the protein/enzyme ratio. Labelled peptides do not balance for this aspect. Second, some other sample components, such as polyphenols and tannins, may also affect the efficiency of trypsin digestion (Gonçalves, Mateus, Pianet, Laguerre, & De Freitas, 2011), which might help to explain why the chocolate matrix gave lower signals for most of targeted peptides. Contrary to the labelled peptides, the U-$^{15}$N concatemer needed to be digested by trypsin to yield peptides that could be detected upon UHPLC-MS/MS analysis. Therefore, factors such as the sample protein content or the presence of tannins should be balanced when using such an internal standard. However, our results indicated that the performance of the concatemer was peptide-dependent. For some peptides, such as TNDNAQISPLAGR from hazelnut Cor a 9 allergen and FFVAPFPEVFGK from milk $\alpha_{S1}$-casein, matrix effects were efficiently balanced with linear regression lines CV below 15% between the matrices. However, for other peptides, such as NVNFDGEILK from egg vitellogenin-1 and TPEVDDEALEK from milk β–lactoglobulin, the associated CVs were much higher (greater than30%).

Matrix effects can also affect analytes by other means. Robustness to food processing was one of the criteria for peptide biomarkers selection (Gavage et al., 2019, 2020; Van Vlierberghe et al., 2020) and can therefore be excluded. Variation in protein extraction can also be excluded as, for all selected proteins, multiple peptides were included in the U–$^{15}$N concatemer, and no general trend of the matrix effect was observed for all the peptides of a given protein. Indeed, if protein extraction of the analyte and/or the internal standard was affected by the matrix, all peptides from a given protein should be equally impacted, which was not observed.

Proteolytic digestion of extracted proteins is a key step in sample preparation and could be a source of the observed variability. Even if the concatemer internal standard needs to be digested to release its constitutive peptides, multiple factors could influence the digestion kinetics. For instance, amino acids surrounding trypsin recognition sites are known to influence the efficiency of peptide bond hydrolysis (Siepen, Keevil, Knight, & Hubbard, 2007). Cleavage sites are described using the nomenclature formulated by Schechter and Berger (Schechter & Berger, 1967), as *P*4-*P*3-*P*2-*P*1-*P*1′-*P*2′-*P*3′, in which cleavage of the peptide bond occurs between *P*1 and *P*1′. Arginine, lysine, and proline in position *P*1′ have, for instance, a negative effect on the digestion efficiency. The acidic amino acids aspartate and glutamate also negatively influence digestion when they are present near the cleavage site. These aspects were taken into account during peptide biomarkers selection, and sequences known to negatively affect trypsin digestion were rejected. However, peptide biomarkers were synthetically stitched together in the concatemer. Considering a given peptide in the concatemer, its cleavage site is surrounded at the *N*-terminal side (*P*4 to *P*1) by amino acids from this peptide but also by amino acids from its neighboring peptide at the C-terminal side (*P*1′ to *P*3′). Consequently, at a local scale, enzymatic digestion of the concatemer only partially reflects digestion of the natural proteins. This difference between natural analytes and concatemers might lead to differences in enzymatic digestion kinetics and could have been a source of the observed variations. A relatively simple solution to overcome this would be the introduction of amino acids between each targeted peptide of the

concatemer. Such introduced amino acids could be the flanking amino acids in the corresponding natural protein sequence, a solution known as a peptide-concatenated standard (PCS) (Kito, Ota, Fujita, & Ito, 2007). However, amino acids surrounding the cleavage site in the three-dimensional structure of the protein might also affect trypsin digestion. Hence, cleavage sites surrounded by acidic amino acids, characterized by a greater average exposed area, are more subject to missed-cleavages.

Besides flanking amino acids, structural parameters also interfere with enzymatic digestion of a protein. According to the work of Hamady and co-workers (Hamady, Cheung, Tufo, & Knight, 2005), secondary protein structures affect trypsin digestion efficiency. Cleavage sites within unstructured domains are more prone to be cleaved incorrectly, whereas cleavage sites in alpha-helices are more favorable. The structures of proteins targeted by the UHPLC-MS/MS method, when available, were analyzed to define whether observed variability among peptides could be linked to findings of Hamady and co-workers or not (Hamady et al., 2005). No general trend emerged from our data, limited to the 19 concatenated peptides. However, three-dimensional and structural aspects could be included in a future peptide biomarker selection, in addition to all other criteria already considered in this study.

## 4. Conclusions

Mass spectrometry-based detection and quantification of food allergens in processed food products remains challenging. Currently, no threshold values for allergen trace-level contamination have been established in European legislation, but these are highly expected by all stakeholders involved in the food chain, from producers to control laboratories, and will require quantitative analysis methods. Quantitative methods based on stable dilution techniques need isotopically labelled internal standards.

Here, we presented and compared the performances of three different types of isotopically labelled internal standards for allergen analysis in processed food products: synthetic peptides, concatemer, and protein. These internal standards were compared through the analysis of three matrix-matched calibration curves (cookie, chocolate, and unbaked lyophilized cookie dough) for four targeted allergens (egg, milk, peanut, and hazelnut). An effective internal standard needs to behave similar to the natural analyte and is therefore identically impacted by matrix effects during sample preparation and UHPLC-MS/MS analysis. As expected from a theoretical point of view, the isotopically labelled protein that was used as an internal standard gave the best results. A constant signal ratio between the analyte and the internal standard peak areas was observed in all matrices tested for the four tryptic peptides generated from the studied protein. However, we need to emphasize that these results only come from one investigated protein, β-lactoglobulin from milk.

Results from our studies using peptides and the concatemer were more equivocal and seemed to be peptide-dependent. For some synthetic peptides or some tryptic peptides from the concatemer, matrix effects during sample preparation and UHPLC-MS/MS analysis could be efficiently countered by the applied internal standards, whereas for other peptides, significant matrix effects were observed. However, the non-inferiority of the results obtained for the tryptic peptides from the concatemer was established, when compared to synthetic peptides. Moreover, the addition of any synthetic peptide in a method represent an additional cost, limiting therefore the number of targeted peptides for routine laboratories. The concatemer production costs are relatively independent of the number of concatenated tryptic peptides. From a rough estimate of ten peptides, the use of a concatemer as internal standard is financially advantageous and supersedes synthetic peptides.

Even though isotopically labelled synthetic peptides are currently the most commonly used internal standard for allergen analysis, they do not exactly reflect the natural situation as they do not need to be subjected to proteolytic digestion, while part of the variability observed in our study could have come from proteolytic digestion. Concatemers clearly need to be digested to release their constituting peptides. However, our data seem to indicate that the digestion of the concatemer could be improved to more efficiently represent analyte protein digestion. In this respect, introducing flanking amino acids between each individual peptide (i.e. the PCS strategy) could be a future asset. Moreover, for our concatemer construct, peptide biomarker selection was mainly focused on robustness to food processing and local sequences, but additional criteria, such as protein structure and the local digestion site environment, could be included in the peptide selection process. Such possible future improvements strongly suggest that isotopically labelled concatemers could represent relevant internal standards, as they overcome limitations of the use of synthetic peptides, while combining advantages of the use of labelled proteins and, further, allowing for multiple allergen quantification by mass spectrometry.

## CRediT authorship contribution statement

**Maxime Gavage:** Writing - original draft, Conceptualization, Investigation, Visualization. **Kaatje Van Vlierberghe:** Writing - review & editing, Investigation. **Christof Van Poucke:** Writing - review & editing, Conceptualization. **Marc De Loose:** Writing - review & editing. **Kris Gevaert:** Writing - review & editing. **Marc Dieu:** Writing - review & editing. **Patsy Renard:** Conceptualization, Writing - review & editing. **Thierry Arnould:** Conceptualization, Writing - review & editing. **Patrice Filee:** Writing - review & editing, Resources. **Nathalie Gillard:** Project administration, Conceptualization, Writing - review & editing, Supervision.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

The research that yielded these results was funded by the Belgian Federal Public Service of Health, Food Chain Safety and Environment through the contract RT 15/10 ALLERSENS. The authors also want to gratefully acknowledge the contributions of the University of Namur Mass Spectrometry platform (MaSUN).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2020.127413.

## References

Ahsan, N., Rao, R. S. P., Gruppuso, P. A., Ramratnam, B., & Salomon, A. R. (2016). Targeted proteomics: Current status and future perspectives for quantification of food allergens. *Journal of Proteomics*. https://doi.org/10.1016/j.jprot.2016.04.018.

Allen, K. J., Remington, B. C., Baumert, J. L., Crevel, R. W. R., Houben, G. F., Brooke-Taylor, S., ... Taylor, S. L. (2014). Allergen reference doses for precautionary labeling (VITAL 2.0): Clinical implications. *Journal of Allergy and Clinical Immunology*.. https://doi.org/10.1016/j.jaci.2013.06.042.

Boo, C. C., Parker, C. H., & Jackson, L. S. (2018). A targeted LC-MS/MS method for the simultaneous detection and quantitation of egg, milk, and peanut allergens in sugar cookies. *Journal of AOAC International*. https://doi.org/10.5740/jaoacint.17-0400.

Boyce, J. A., Assa'ad, A., Burks, A. W., Jones, S. M., Sampson, H. A., Wood, R. A., ... Schwaninger, J. M. (2011). Guidelines for the Diagnosis and Management of Food Allergy in the United States: Summary of the NIAID-Sponsored Expert Panel Report. *Journal of the American Academy of Dermatology*. https://doi.org/10.1016/j.jaad.2010.11.020.

Chambers, S. P., Austen, D. A., Fulghum, J. R., & Kim, W. M. (2004). High-throughput screening for soluble recombinant expressed kinases in Escherichia coli and insect cells. *Protein Expression and Purification*. https://doi.org/10.1016/j.pep.2004.03.003.

Galan, S. R. G., Wickens, J. R., Dadova, J., Ng, W. L., Zhang, X., Simion, R. A., ... Davis, B.

G. (2018). Post-translational site-selective protein backbone α-deuteration. *Nature Chemical Biology*. https://doi.org/10.1038/s41589-018-0128-y.

Gavage, M., Van Vlierberghe, K., Van Poucke, C., De Loose, M., Gevaert, K., Dieu, M., ... Gillard, N. (2019). Selection of egg peptide biomarkers in processed food products by high resolution mass spectrometry. *Journal of Chromatography A, 1584*, 115–125. https://doi.org/10.1016/J.CHROMA.2018.11.036.

Gavage, M., Van Vlierberghe, K., Van Poucke, C., De Loose, M., Gevaert, K., Dieu, M., ... Gillard, N. (2020). High-resolution mass spectrometry-based selection of peanut peptide biomarkers considering food processing and market type variation. *Food Chemistry*. https://doi.org/10.1016/j.foodchem.2019.125428.

Georgiou, C. A., & Danezis, G. P. (2015). Elemental and Isotopic Mass Spectrometry. *Comprehensive Analytical Chemistry, 68*, 131–243. https://doi.org/10.1016/B978-0-444-63340-8.00003-0.

Gonçalves, R., Mateus, N., Pianet, I., Laguerre, M., & De Freitas, V. (2011). Mechanisms of tannin-induced trypsin inhibition: a molecular approach. *Langmuir*. https://doi.org/10.1021/la202280c.

Gorochowski, T. E., Ignatova, Z., Bovenberg, R. A. L., & Roubos, J. A. (2015). Trade-offs between tRNA abundance and mRNA secondary structure support smoothing of translation elongation rate. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkv199.

Hamady, M., Cheung, T. H. T., Tufo, H., & Knight, R. (2005). Does protein structure influence trypsin miscleavage? *IEEE Engineering in Medicine and Biology Magazine, 24*(3), 58–66. https://doi.org/10.1109/MEMB.2005.1436461.

Henrottin, J., Planque, M., Huet, A.-C., Marega, R., Lamote, A., & Gillard, N. (2019). Gluten analysis in processed foodstuffs by a multi-allergens and grain-specific UHPLC-MS/MS method: One method to detect them all. *Journal of AOAC International*. https://doi.org/10.5740/jaoacint.19-0057.

Kito, K., Ota, K., Fujita, T., & Ito, T. (2007). A synthetic protein approach toward accurate mass spectrometric quantification of component stoichiometry of multiprotein complexes. *Journal of Proteome Research*. https://doi.org/10.1021/pr060447s.

Korte, R., Oberleitner, D., & Brockmeyer, J. (2019). Determination of food allergens by LC-MS: Impacts of sample preparation, food matrix, and thermal processing on peptide detectability and quantification. *Journal of Proteomics, 196*, 131–140. https://doi.org/10.1016/J.JPROT.2018.11.002.

Loch, J. I., Bonarek, P., Tworzydło, M., Polit, A., Hawro, B., Łach, A., ... Lewiński, K. (2016). Engineered β-Lactoglobulin Produced in E. coli: Purification, Biophysical and Structural Characterisation. *Molecular Biotechnology.*. https://doi.org/10.1007/s12033-016-9960-z.

MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., ... MacCoss, M. J. (2010). Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btq054.

Monaci, L., Losito, I., De Angelis, E., Pilolli, R., & Visconti, A. (2013). Multi-allergen quantification of fining-related egg and milk proteins in white wines by high-resolution mass spectrometry. *Rapid Communications in Mass Spectrometry, 27*(17), 2009–2018. https://doi.org/10.1002/rcm.6662.

Mondal, S., Shet, D., Prasanna, C., & Atreya, H. S. (2013). High yield expression of proteins in E. coli for NMR studies. Advances in Bioscience and Biotechnology. https://doi.org/10.4236/abb.2013.46099.

Newsome, G. A., & Scholl, P. F. (2013). Quantification of allergenic bovine milk αs1-casein in baked goods using an intact 15N-labeled protein internal standard. *Journal of Agricultural and Food Chemistry*. https://doi.org/10.1021/jf3015238.

Nitride, C., Nørgaard, J., Omar, J., Emons, H., Esteso, M. J. M., & O'Connor, G. (2019). An assessment of the impact of extraction and digestion protocols on multiplexed targeted protein quantification by mass spectrometry for egg and milk allergens. *Analytical and Bioanalytical Chemistry*. https://doi.org/10.1007/s00216-019-01816-z.

Parker, C. H., Khuda, S. E., Pereira, M., Ross, M. M., Fu, T. J., Fan, X., ... Jackson, L. S. (2015). Multi-allergen Quantitation and the Impact of Thermal Treatment in Industry-Processed Baked Goods by ELISA and Liquid Chromatography-Tandem Mass Spectrometry. *Journal of Agricultural and Food Chemistry*. https://doi.org/10.1021/acs.jafc.5b04287.

Planque, Arnould, Delahaut, Renard, Dieu, & Gillard. (2019). Development of a strategy for the quantification of food allergens in several food products by mass spectrometry in a routine laboratory. Food Chemistry. https://doi.org/10.1016/j.foodchem.2018.08.095.

Planque, Arnould, & Gillard. (2017). Food Allergen Analysis: Detection, Quantification and Validation by Mass Spectrometry. In Allergen. https://doi.org/10.5772/intechopen.69361.

Planque, M., Arnould, T., Dieu, M., Delahaut, P., Renard, P., & Gillard, N. (2016). Advances in ultra-high performance liquid chromatography coupled to tandem mass spectrometry for sensitive detection of several food allergens in complex and processed foodstuffs. Journal of Chromatography A, 1464, 115–123. https://doi.org/http://doi.org/10.1016/j.chroma.2016.08.033.

Pratt, J. M., Simpson, D. M., Doherty, M. K., Rivers, J., Gaskell, S. J., & Beynon, R. J. (2006). Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nature Protocols*. https://doi.org/10.1038/nprot.2006.129.

Schechter, I., & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochemical and Biophysical Research Communications, 27*(2), 157–162. https://doi.org/10.1016/S0006-291X(67)80055-X.

Sicherer, S. H., & Sampson, H. A. (2014). Food allergy: Epidemiology, pathogenesis, diagnosis, and treatment. *Journal of Allergy and Clinical Immunology*. https://doi.org/10.1016/j.jaci.2013.11.020.

Siepen, J. A., Keevil, E. J., Knight, D., & Hubbard, S. J. (2007). Prediction of missed cleavage sites in tryptic peptides aids protein identification in proteomics. *Journal of Proteome Research, 6*(1), 399–408. https://doi.org/10.1021/pr060507u.

Van Vlierberghe, K., Gavage, M., Dieu, M., Renard, P., Arnould, T., Gillard, N., ... Van Poucke, C. (2020). Selection of universal peptide biomarkers for the detection of the allergen hazelnut in food trough a comprehensive, high resolution mass spectrometric (HRMS) based approach. *Food Chemistry, 125679*. https://doi.org/10.1016/J.FOODCHEM.2019.125679.

## 5. Conclusion

The combination of targeted proteomics and isotope dilution are widely described for protein quantification (Villanueva *et al*, 2014). However, the choice of the internal standard is crucial to ensure method performance. For food allergen analysis, isotope-labelled protein is the ideal internal standard counterbalancing the analyte loss and matrix effects at all the steps of the sample preparation and during the analysis method. Neglecting technical issues, this approach is currently unaffordable in routine analysis. Therefore, most laboratories rely on isotope-labelled synthetic peptides. However, peptides do not undergo the enzymatic digestion, a step which is known to be affected by the food matrix.

Here, an alternative method was developed, based on the use of an isotope-labelled concatemer as internal standard for allergen quantification. This technique is known for more than a decade in the proteomic area but was never applied to food allergen quantification nor even in food analysis. Concatemers need to be proteolytically digested and are also affected by the interference caused by the matrix. This approach has also the advantage of allowing multiplex analysis.

Several constructs were developed to identify critical design criteria and production parameters. A $^{15}N$ isotope-labelled concatemer composed of 19 peptide biomarkers from four allergenic ingredients was finally synthetized. In addition, the production yield of the concatemers obtained by this approach is cost-effective when compared to the use of synthetic peptides as internal standards. Isotope labelling strategy and obtained isotopic enrichment were adequate to avoid the introduction of any risk for false positive identification.

The performance of the synthetized concatemer in the quantitative analysis was finally compared to the one of a labelled protein and synthetic peptides in the analysis of food matrices spiked with food allergen extracts. As expected from a theoretical point of view, the isotopically labelled protein that was used as an internal standard gave the best results. Synthetic peptides and the concatemer were more equivocal and seemed to be peptide-dependent. Several possibilities of improvement were proposed, such as the addition of flanking sequences between each peptide biomarker to more efficiently represent analyte protein digestion. Such possibilities strongly suggest that isotopically labelled concatemers could represent relevant internal standards, as they overcome limitations of the use of synthetic peptides, while combining advantages of the use of labelled proteins and, further, allowing for multiple allergen quantification by mass spectrometry.

**PART III**

## Development and validation of the UHPLC-MS/MS Method for the detection and quantification of 4 allergenic ingredients (egg, milk, peanut and hazelnut) in processed food products

### 1. Introduction

The last step in the development of a routine analysis method for the detection and quantification of food allergens in processed food matrices was the development and validation of the UHPLC-MS/MS method. A series of potential peptide biomarkers were previously identified through HRMS analysis of various processed test matrices separately containing egg, milk, peanut and hazelnut. These potential peptide biomarkers were used to develop the UHPLC-MS/MS method. The number of peptide biomarkers was reduced to only keep the best performers in terms of detection sensitivity and selectivity. This task was entirely realized by Kaatje Van Vlierberghe at ILVO.

To develop and validate the UHPLC-MS/MS method a second type of test material was produced. These food matrices were incurred with trace levels (0.5 ppm to 25 ppm) of the four allergenic ingredients. The term incurred means that the allergenic ingredients were mixed with the other ingredients of the matrix before their processing (such as cooking). The homogeneous distribution of the allergenic ingredients in the incurred matrices was a critical aspect for the method validation. This matrix homogeneity was evaluated with ELISA analysis.

As we have seen, a $^{15}N$ stable isotope-labelled concatemer was developed to be used as an internal standard in the developed UHPLC-MS/MS method. This concatemer was composed of the final 19 peptide biomarkers selected during the UHPLC-MS/MS method development.

The developed UHPLC-MS/MS method was finally validated combining the use of the $^{15}N$ stable isotope-labelled concatemer as internal standard and standard addition method for quantification. Performance parameters including selectivity, LOD, LOQ, linearity, trueness and precision were evaluated. These results were compared to standard method performance required by the AOAC for the detection and quantification of food allergens. These requirements were already addressed and presented in Table 3 of the introduction chapter.

### 2. Test material production (done entirely by Kaatje Van Vlierberghe at ILVO)

The second type of test material contained all four allergenic ingredients together. The food matrices and processing conditions were selected as a function of representativeness for real food products and of processing conditions that are applied nowadays in the food industry. Food matrices and processing conditions were also selected according to their well-known effect on allergen detectability, such as heat effect or the presence of tannins. Test material were prepared by Kaatje Van Vlierberghe in the Food Pilot unit of ILVO.

These allergen incurred food matrices were used to develop and validate the UHPLC-MS/MS method and will also be used by Kaatje Van Vlierberghe to compare the developed method with other existing allergen analysis methods (WP5 of the “Allersens” project).

Initially, three food matrices were selected and produced with allergenic ingredients (egg, milk, peanut and hazelnut) incurred at different levels (0, 0.5, 1, 2.5, 5, 10, 25 ppm expressed in mg total allergen protein per kg of matrix). These three matrices were:

- Rice powder as a neutral background matrix
- Cookie to include heat-process impact (25 min at 180 °C)
- Chocolate dessert as fat and complex matrix containing tannin. This matrix was prepared in a powder form and had to be mixed with water to obtain the true chocolate dessert matrix.

However, homogeneity testing (extensively detailed in the next section) of these three first matrices indicated that allergenic ingredients were not homogeneously distributed in rice powder and chocolate dessert even after multiple attempts. These two matrices were therefore replaced by:

- Unbaked freeze-dried cookie dough as a neutral background matrix
- Chocolate chip as fat and complex matrix containing tannin.

To include potential variability in protein content and protein distribution due to growing conditions and origin, peanut and hazelnut from two distinct geographical growing locations were considered for the production of test materials. Therefore, each combination of food matrix and contamination level was prepared in duplicate. Type 1 matrices corresponded to peanut from China and hazelnut from Turkey, whereas type 2 matrices corresponded to peanut from Israel and hazelnut from Italy.

### 3. Test material homogeneity testing (done entirely by Maxime Gavage at CER Groupe)

Homogeneity of incurred test matrices was evaluated using direct sandwich ELISA kits developed at CER Groupe. The UHPLC-MS/MS could not be used before its validation and this validation required homogeneous test matrices. The statistical approach developed by Fearn and Thompson (Fearn & Thompson, 2001) was employed to assess the sufficient homogeneity of test materials. That is, the variance in the mean composition of the distributed portions of the material must be negligibly small in relation to the variance of the analytical result produced when the material is in normal use. The statistical test is therefore based on the comparison of the sampling variance with a target variance based on the Horwitz equation, defining a relationship between the precision of an analytical method and the concentration of the analyte (Horwitz *et al*, 1980).

For each matrix contamination level (0.5, 1, 2.5, 5, 10 and 25 ppm expressed in mg total allergen protein per kg of matrix), matrix (rice flour, cookie and chocolate dessert and later unbaked freeze-dried cookie dough and chocolate chip) and each matrix type (for the two peanut and hazelnut origins), 10 samples of 2 g were collected in 50 ml conical tubes (for chocolate dessert, 0.67 g were collected and resuspended with 1.33 ml of water to obtain 2 g of matrix). A volume of 20 ml of pre-heated (at 60°C) extraction buffer (0.05 M PBS with 1 % Tween 20 (10 ml/l) and 1.5 % gelatine (15 g/L)) is added to each sample. Extraction was conducted for 30 min in a 60°C water bath under 160 rpm orbital agitation. Samples were then centrifuged (4660 g for 15 min) and the resulting supernatant was collected in a new 50 ml conical tube. In parallel, for each matrix, a calibration curve was established using a blank matrix (uncontaminated with the four allergenic ingredients) spiked with defined quantities of extracts of the four considered allergenic ingredients (egg, milk, peanut and hazelnut).

Each sample extract was analysed in duplicate, for the four allergenic ingredients, on randomly allocated 96-microwell plates. Three incubation periods of 30 min, separated by a washing step, were carried out (sample reaction with capture antibody, reaction of the capture antibody–antigen complex with enzyme-conjugated antibody and reaction between enzyme and substrate). Allergen

concentration was finally measured based on colorimetric measurements referred to a calibration curve.

An example of obtained results is exposed in Figure 60. The limit of quantification of each ELISA kit test selects the first point of the calibration curve. This point is kit specific (0.25 ppm for egg, 0.5 ppm for milk, 0.25 ppm for peanut and 0.5 ppm for hazelnut, ppm expressed in mg total allergen protein per kg of matrix). A measured allergen concentration was accepted only if it is higher than this LOQ.

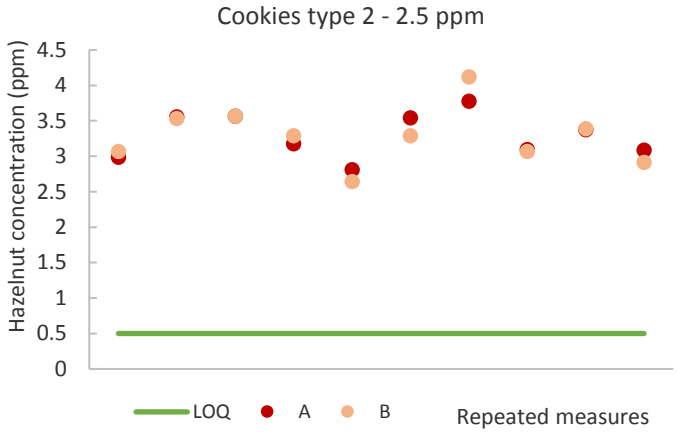



*Figure 60 - Measured hazelnut concentration for cookie type 2 contaminated at 2.5 ppm. Ten independent replicates were prepared and analysed in duplicate with direct sandwich ELISA kits developed at CER Groupe.*

For each combination of matrix, matrix type, allergenic ingredient and contamination level (240 tests), allergenic ingredient distribution homogeneity was evaluated using the Fear and Thompson test (Fearn & Thompson, 2001). Results of this homogeneity test for rice flour, cookie and chocolate dessert are presented in Table 9. As complementary information, average measured allergenic ingredient concentration (in ppm) and relative standard deviation (RSD) are detailed.

Obtained homogeneity testing results indicated that cookies were homogeneous for 3 allergenic ingredients (peanut, milk and hazelnut) at all contamination levels (except 0.5 ppm for which measured concentrations were below ELISA kits LOQ). Egg was not detectable in cookies because of thermal processing, known to affect allergen detection (Gomaa & Boye, 2013). Therefore, no conclusion could be drawn regarding homogeneity.

On the opposite, allergenic ingredients were found to be globally not homogeneously distributed in chocolate dessert matrix. Milk distribution seemed to be better when compared to other allergenic ingredients. This was most likely due to the higher variability between duplicates of a same sample, inducing a higher tolerance to the test.

Concerning rice flour matrix, no general trends were discerned. Allergenic ingredients were homogeneously distributed for certain contamination levels but not for others. It was however observed that egg was generally not homogeneously distributed.

In certain conditions, measured allergenic ingredient concentration was over-estimated, mostly for high concentrations. This observation could be explained by the fact that concentrated extracts needed to be diluted to be included in the calibration curve. Matrix components are diluted as well, reducing associated matrix effects, which was not the case for the calibration curve.

*Table 9 - Summarized results of homogeneity evaluation using the Fearn and Thompson test for rice flour, cookie and chocolate dessert matrices. Homogeneity was indicated using a colour code: green for homogeneous, red for not homogeneous and yellow for test invalidity because of measured concentrations below ELISA kit LOQ. Concentration and RSD results are expressed as the mean value of the 10 replicates (analysed in duplicate) of each matrix and each contamination level.*

| | | Rice flour type 1 | | | | | | Rice flour type 2 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0.5 ppm | 1 ppm | 2.5 ppm | 5 ppm | 10 ppm | 25 ppm | 0.5 ppm | 1 ppm | 2.5 ppm | 5 ppm | 10 ppm | 25 ppm |
| Peanut | Concentration (ppm) | 0.47 | 0.73 | 3.38 | 6.03 | 13.56 | 29.53 | 0.49 | 0.83 | 4.01 | 6.40 | 14.96 | 30.93 |
| | RSD (%) | 2.72 | 21.25 | 23.97 | 10.42 | 8.48 | 4.92 | 9.21 | 41.01 | 59.20 | 13.63 | 15.69 | 8.09 |
| Egg | Concentration (ppm) | < LOQ | 0.94 | 2.48 | 3.43 | 9.54 | 21.48 | < LOQ | 0.77 | 1.42 | 4.53 | 9.30 | 20.11 |
| | RSD (%) | < LOQ | 55.84 | 56.48 | 17.59 | 26.96 | 17.46 | < LOQ | 48.21 | 18.41 | 53.93 | 26.19 | 12.05 |
| Milk | Concentration (ppm) | < LOQ | 0.90 | 3.46 | 6.91 | 19.74 | 45.89 | < LOQ | 1.10 | 2.71 | 7.13 | 18.30 | 48.18 |
| | RSD (%) | < LOQ | 14.99 | 30.77 | 12.59 | 20.11 | 7.84 | < LOQ | 29.82 | 14.34 | 35.18 | 11.57 | 9.98 |
| Hazelnut | Concentration (ppm) | 1.62 | 2.06 | 3.94 | 8.20 | 24.14 | 58.37 | 1.59 | 2.29 | 3.62 | 7.68 | 23.82 | 57.83 |
| | RSD (%) | 15.62 | 8.17 | 7.19 | 11.36 | 5.82 | 3.80 | 17.77 | 35.37 | 8.41 | 5.16 | 6.05 | 3.23 |

| | | Cookie type 1 | | | | | | Cookie type 2 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0.5 ppm | 1 ppm | 2.5 ppm | 5 ppm | 10 ppm | 25 ppm | 0.5 ppm | 1 ppm | 2.5 ppm | 5 ppm | 10 ppm | 25 ppm |
| Peanut | Concentration (ppm) | < LOQ | 0.62 | 1.63 | 3.83 | 6.9 | 19.06 | < LOQ | 0.55 | 1.48 | 3.54 | 6.6 | 18.62 |
| | RSD (%) | < LOQ | 19.34 | 12.69 | 11.72 | 7.86 | 10.13 | < LOQ | 14.45 | 12.3 | 5.62 | 5.35 | 8.61 |
| Egg | Concentration (ppm) | < LOQ | < LOQ | < LOQ | < LOQ | < LOQ | < LOQ | < LOQ | < LOQ | < LOQ | < LOQ | < LOQ | < LOQ |
| | RSD (%) | < LOQ | < LOQ | < LOQ | < LOQ | < LOQ | < LOQ | < LOQ | < LOQ | < LOQ | < LOQ | < LOQ | < LOQ |
| Milk | Concentration (ppm) | < LOQ | 0.34 | 0.56 | 1.35 | 2.6 | 9.36 | < LOQ | 0.43 | 0.8 | 1.77 | 3.53 | 12.98 |
| | RSD (%) | < LOQ | 22.76 | 27.88 | 14.83 | 19.55 | 15.29 | < LOQ | 25.91 | 17.91 | 17.58 | 15.39 | 20.18 |
| Hazelnut | Concentration (ppm) | < LOQ | 1.92 | 3.29 | 8.22 | 16.11 | 40.51 | < LOQ | 1.85 | 3.29 | 7.9 | 15.51 | 41.78 |
| | RSD (%) | < LOQ | 10.26 | 13.45 | 9.25 | 8.44 | 9.68 | < LOQ | 7.05 | 10.71 | 7.79 | 6.32 | 7.32 |

| | | Chocolate dessert type 1 | | | | | | Chocolate dessert type 2 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0.5 ppm | 1 ppm | 2.5 ppm | 5 ppm | 10 ppm | 25 ppm | 0.5 ppm | 1 ppm | 2.5 ppm | 5 ppm | 10 ppm | 25 ppm |
| Peanut | Concentration (ppm) | < LOQ | 1.17 | 4.11 | 9.32 | 34.72 | 80.97 | < LOQ | 2.02 | 5.22 | 12.58 | 25.4 | 92.38 |
| | RSD (%) | < LOQ | 63.42 | 25.1 | 29.48 | 84.37 | 15.47 | < LOQ | 71.37 | 18.81 | 21.13 | 22.6 | 23.65 |
| Egg | Concentration (ppm) | < LOQ | 0.83 | 2.32 | 6.03 | 9.99 | 22.44 | < LOQ | 0.42 | 3.24 | 6.84 | 11.04 | 23.6 |
| | RSD (%) | < LOQ | 134.63 | 93.6 | 68.22 | 57.45 | 39.49 | < LOQ | 100.11 | 72.17 | 78.28 | 53.05 | 31.7 |
| Milk | Concentration (ppm) | < LOQ | 0.98 | 2.62 | 4.94 | 13.25 | 38.11 | < LOQ | 1.06 | 2.83 | 5.62 | 13.02 | 37.97 |
| | RSD (%) | < LOQ | 44.9 | 37.66 | 25.94 | 35.15 | 32.57 | < LOQ | 23.88 | 31.88 | 31.6 | 34.02 | 23.28 |
| Hazelnut | Concentration (ppm) | < LOQ | 2.07 | 4.07 | 10.76 | 23.83 | 76.94 | < LOQ | 1.84 | 4.64 | 13.21 | 22.37 | 75.28 |
| | RSD (%) | < LOQ | 11.11 | 21.21 | 26.88 | 16.93 | 29.75 | < LOQ | 12.05 | 31.17 | 32.95 | 19.94 | 22.13 |

Sample weighing of chocolate dessert was slightly different compared to other matrices. The contamination level was given for the reconstituted product (1/3 powder and 2/3 water). In the performed analysis, 0.67 g of powder was weighed and reconstituted with 1.33 ml of water to finally obtain 2 g of reconstituted product. The amount of matrix weighed was reduced compared to other matrices and this could impact homogeneity testing result (Pauwels *et al*, 1998).

To investigate the effects of weighed amounts, additional analyses were performed on chocolate dessert with two different weighing techniques.

The first weighing technique simply corresponded to a weighing of 2 g of powder without any reconstitution with water. The powder was directly extracted with 20 ml of extraction buffer. The targeted contamination level was no longer applicable but homogeneity testing could still be performed. This procedure was applied on the 10 samples of chocolate type 1 contaminated at 5 ppm level. Homogeneity testing was performed for all 4 allergenic ingredients.

For the second weighing technique, chocolate dessert was reconstituted as previously. In a 50 ml conical tube, 10 g of powder was mixed with 20 ml of water using vortex agitation. Ten samples of 2 g were subsequently collected and analysed.

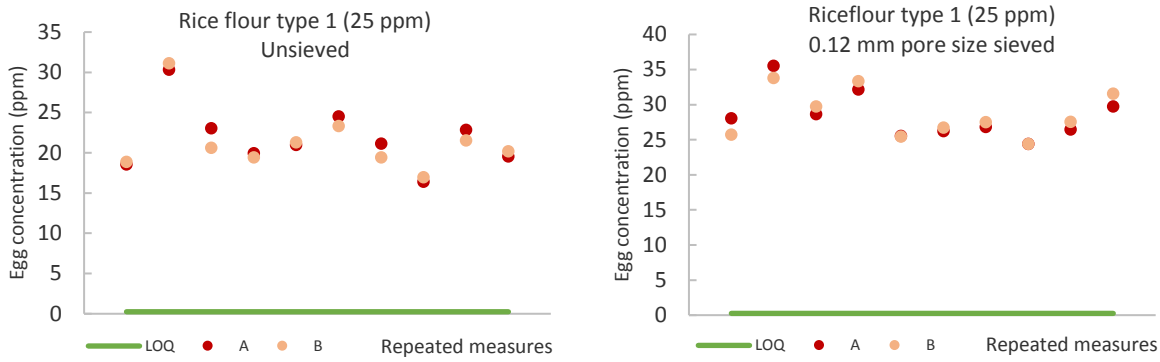
*Table 10 – Impact of chocolate dessert weighing technique on allergenic ingredient distribution homogeneity evaluated with the Fearn and Thompson test on the 5 ppm contamination level. Homogeneity is indicated using a colour code: green for homogeneous, red for not homogeneous. Concentration and RSD results are expressed as the mean value of the 10 replicates (analysed in duplicate) of each weighing technique.*

| | | Chocolate dessert type 2 (5 ppm) | | |
|---|---|---|---|---|
| | | 0.67 g of powder + 1.33 ml of water | 2 g of powder without reconstitution | 2 g of previously reconstituted product |
| Peanut | Concentration (ppm) | 12.58 | 22.51 | 10.50 |
| | RSD (%) | 21.13 | 41.96 | 23.62 |
| Egg | Concentration (ppm) | 6.84 | 16.26 | 3.93 |
| | RSD (%) | 78.28 | 46.40 | 29.52 |
| Milk | Concentration (ppm) | 5.62 | 12.55 | 6.66 |
| | RSD (%) | 31.6 | 28.96 | 22.30 |
| Hazelnut | Concentration (ppm) | 13.21 | 44.86 | 12.05 |
| | RSD (%) | 32.95 | 25.18 | 24.39 |

Results of the evaluation of the weighing technique on allergenic ingredient distribution are exposed in Table 10. Obtained results indicated that, independently of the sample weighing technique, allergenic ingredients were not homogeneously distributed in the food matrix. Results were obtained for the type 2 product at a 5 ppm contamination level but could certainly be extended to other contamination levels.

Another investigated hypothesis to explain the insufficient homogeneity found for chocolate dessert and rice flour was the particle size distribution. Particle size is known to affect mixture quality of powders (Shenoy *et al*, 2015). To test this hypothesis, rice flour and allergenic ingredients were sieved with the same sieve (0.12 mm pore size) for the preparation of the 25 ppm contamination level.

Obtained results (Figure 61) indicated that a sufficient homogeneity was not obtained for egg despite the use of a 0.12 mm pore size sieve. A visual improvement was however observed, with lower spreading of measured concentration values, also confirmed by a lower RSD.




| | | Rice flour type 1 (25 ppm) | |
|---|---|---|---|
| | | Unsieved | 0.12 mm pore size sieved |
| **Peanut** | Concentration (ppm) | 29.53 | 32.68 |
| | RSD (%) | 4.92 | 7.1 |
| **Egg** | Concentration (ppm) | 21.48 | 28.43 |
| | RSD (%) | 17.46 | 11.5 |
| **Milk** | Concentration (ppm) | 45.89 | 18.79 |
| | RSD (%) | 7.84 | 12.1 |
| **Hazelnut** | Concentration (ppm) | 58.37 | 62.32 |
| | RSD (%) | 3.80 | 2.6 |

*Figure 61 - Impact of food matrix and allergenic ingredient sieving on allergenic ingredient distribution homogeneity evaluated with the Fearn and Thompson test on the 25 ppm contamination level. Homogeneity is indicated using a colour code: green for homogeneous, red for not homogeneous. Concentration and RSD results are expressed as the mean value of the 10 replicates (analysed in duplicate) of each matrix (unsieved and 0.12 mm pore size sieved).*

Difficulties to obtain homogeneous matrices could be due to powder mixing and particle size distribution. To circumvent this problem, powder matrices were replaced by matrices including a liquid state in their preparation (like the dough in the cookie matrix). Chocolate dessert powder was replaced by chocolate chips and rice flour by freeze-dried cookie dough. Allergenic ingredients were incorporated and mixed with matrices under a liquid phase (melted chocolate and cookie dough before freeze-drying).

In order to save time, only type 2 chocolate chips and freeze-dried cookie dough were prepared. Influence of peanut and hazelnut geographical origin was investigated in HRMS biomarkers identification. Peptide biomarkers finally selected for the UHPLC-MS/MS quantitative methods were derived from proteins observed at comparable level regardless of the geographical origin.

Obtained results (Table 11) indicated that a sufficiently homogeneous allergenic ingredient distribution in chocolate chips was achieved for all four allergenic ingredient and for all contamination levels. According to Fearn and Thompson test, sufficient homogeneity was not achieved in a limited number of cases for the freeze-dried cookie dough. However, the obtained RSD were relatively low and the matrix was conserved.

*Table 11 – Summarized results of homogeneity evaluation using the Fearn and Thompson test for chocolate chips and freeze-dried cookie dough matrices. Homogeneity was indicated using a colour code: green for homogeneous, red for not homogeneous and yellow for test invalidity because of measured concentrations below ELISA kit LOQ. Concentration and RSD results are expressed as the mean value of the 10 replicates (analysed in duplicate) of each matrix.*

| | | Chocolate chips | | | | | |
|---|---|---|---|---|---|---|---|
| | | 0.5 ppm | 1 ppm | 2.5 ppm | 5 ppm | 10 ppm | 25 ppm |
| Peanut | Concentration (ppm) | 1.09 | 2.4 | 6.86 | 14.51 | 22.15 | 54.27 |
| | RSD (%) | 4.8 | 8.3 | 5.6 | 4.6 | 6 | 7.4 |
| Egg | Concentration (ppm) | 0.45 | 0.94 | 2.65 | 5.03 | 10.98 | 27.68 |
| | RSD (%) | 12.2 | 11 | 14.1 | 12.5 | 9.1 | 8.7 |
| Milk | Concentration (ppm) | 0.68 | 1.68 | 2.68 | 5.84 | 16.43 | 45.05 |
| | RSD (%) | 14.3 | 21.9 | 22.9 | 23.1 | 17.5 | 19.8 |
| Hazelnut | Concentration (ppm) | 1.65 | 2.34 | 6.5 | 15.24 | 34.16 | 94.48 |
| | RSD (%) | 1.9 | 4.9 | 4.9 | 5.1 | 6 | 3.7 |
| | | Freeze-dried cookie dough | | | | | |
| | | 0.5 ppm | 1 ppm | 2.5 ppm | 5 ppm | 10 ppm | 25 ppm |
| Peanut | Concentration (ppm) | 0.44 | 0.92 | 2.47 | 4.7 | 22.65 | 56.83 |
| | RSD (%) | 24.5 | 15.8 | 8.3 | 9.4 | 13.2 | 16.6 |
| Egg | Concentration (ppm) | 0.46 | 0.47 | 1.02 | 4.5 | 10.46 | 26.59 |
| | RSD (%) | 27.8 | 17.7 | 9.8 | 10.5 | 15.1 | 11.5 |
| Milk | Concentration (ppm) | < LOQ | 0.79 | 1.11 | 4.22 | 58.14 | 79.95 |
| | RSD (%) | < LOQ | 13.6 | 21.1 | 23 | 20.5 | 23.1 |
| Hazelnut | Concentration (ppm) | 1.62 | 2.24 | 5.5 | 11.78 | 25.3 | 63.78 |
| | RSD (%) | 12.3 | 10.6 | 8.6 | 6.8 | 6.9 | 6.7 |

Cookie, chocolate chips and freeze-dried cookie dough incurred matrices were used to develop the UHPLC-MS/MS based on the list of potential peptide biomarkers identified by HRMS. These matrices were also used to evaluate the method limit of detection (LOD) and the limit of quantification (LOQ) for each one of the 19 final peptide biomarkers.

The uncontaminated version of these matrices was spiked with allergen extracts to evaluate method linearity, trueness and precision. With this approach, only the method variability was considered and not the variability of allergen content.

## 4. Development of the UHPLC-MS/MS method

### 4.1. Sample preparation

Sample preparation for UHPLC-MS/MS analysis was based on the work of Mélanie Planque in the "Allermass" project (Planque *et al*, 2019). The internal standard approach was different with the use of the designed $^{15}N$ isotope-labelled concatemer.

Briefly a volume of 100 µl corresponding to 8 µg of concatemer (12.5x dilution of the 1 mg/ml stock solution) is added to each 2 g sample. Proteins were extracted in 50 mL conical tubes with 20 mL of extraction buffer (200 mM Tris, pH 9.2, 2 M urea) by shaking at 20 °C for 30 min prior to ultrasonic treatment at 4 °C for 15 min. After centrifugation (4660 g, 10 min), 10 mL of supernatant were diluted in digestion buffer (200 mM ammonium bicarbonate, pH 8.2). Protein disulfide bridges were successively reduced and alkylated with 45 min incubation steps at room temperature with the addition of 1 mL of 200 mM dithiothreitol and 1 mL of 400 mM iodoacetamide (in the dark). Proteins

were then enzymatically digested with the addition of 1 mL of trypsin solution (trypsin from bovine pancreas, 1 mg/mL in 50 mM acetic acid, pH 2.8) and incubation for 1 h at 37 °C. The digestion reaction was stopped by adding 300 µl of 20 % formic acid (FA) to the samples, which were then centrifuged (4660 g, 5 min).

Obtained peptides were then purified and concentrated using C18 SPE cartridges, which were first conditioned with 18 mL of acetonitrile (ACN) followed by 18 mL of 0.1 % FA before loading of 20 mL of the centrifuged sample. The cartridges were washed with 18 mL of 0.1 % FA and eluted in 15 mL conical tubes with 6 mL of 80 % ACN and 0.1 % FA. A volume of 30 µl of dimethyl sulfoxide was added to the sample before evaporation (40 °C under nitrogen flow) to avoid dryness. The pellet was finally dissolved in 600 µl of 5 % ACN with 0.1% FA and centrifuged twice (4660 g for 5 min in conical tube and 20 000 g for 5 min in 1.5 mL microtube, keeping the supernatant) before UHPLC-MS/MS analysis.

### 4.2. UHPLC-MS/MS method parameters

Prepared samples were analysed by ultra-high-performance liquid chromatography–tandem mass-spectrometry (UHPL-MS/MS) using an Acquity liquid chromatograph equipped with a C18 Acquity BEH130 column (2.1 x 150 mm; 1.7 µm) and coupled with a Xevo TQ-S triple quadrupole system (Waters).

The peptides were separated by reverse-phase chromatography on-line connected to a triple quadrupole mass spectrometer. The following 26 min solvent gradient (solvent A, 0.1 % FA and solvent B, ACN and 0.1 % FA) was applied to the 20 µl injected sample volume: 0–3 min: 92 % solvent A; 3–18 min: linear gradient from 92 % to 58 % solvent A; 18–22.5 min: 15 % solvent A; and 22.5–26 min: 92 % solvent A, always at constant 0.2 mL/min flow rate. Eluted peptides were ionized using the positive electrospray source and analysed in MRM mode. The source gas flow was set at 50 L/h and the source voltage at 2.5 kV for the capillary and 30 V for the cone. The source temperature was set at 150 °C and the “desolvation” temperature at 400 °C with a gas flow at 1200 L/h.

### 4.3. Selection of the final peptide biomarkers and development of the MRM method

The selection of the final peptide biomarkers for UHPLC-MS/MS method was performed by Kaatje Van Vlierberghe at ILVO during the development of the $^{15}N$ isotope-labelled concatemer internal standard. A summary of obtained results is presented.

Thanks to HRMS analysis, 55 peptides were identified as potential peptide biomarkers for the quantification of four allergenic ingredients (egg, milk, peanut and hazelnut) in processed food products. This list was reduced to only keep the best performers in terms of detection sensitivity and selectivity.

The first step was the development of MRM methods with the most intense transitions for each one of these 55 peptides. To do so, unprocessed ingredients were first analysed with HRMS to identify the most intense precursor ion(s) (the charge state(s) with the highest signal) for each peptide. For these precursor ions, MRM methods including all potential transitions were developed using the open source Skyline software (MacLean *et al*, 2010). Unprocessed allergenic ingredients were analysed with the triple quadrupole mass spectrometer using these MRM methods. For each peptide, the 3 transitions with the highest signal were selected to form a MRM method for each peptide.

Test material (cookie, chocolate chips and freeze-dried cookie dough) incurred with traces levels of the four allergenic ingredients were analysed with the developed MRM methods. Peptides allowing the

detection of allergenic ingredients at the lowest level were selected. Signal selectivity was evaluated, for the natural peptide and for its $^{15}N$ isotope-labelled version, with the analysis of uncontaminated matrices.

A list of 19 peptides biomarkers (Table 12) was finally retained for the quantification of the four allergenic ingredients with UHPLC-MS/MS.

*Table 12 – Final list of 19 peptide biomarkers for the quantification of four allergenic ingredient in processed food products by UHPLC-MS/MS*

| | | |
|---|---|---|
| Milk | $\alpha_{s1}$-casein | FFVAPFPEVFGK, HQGLPQEVLNENLLR, YLGYLEQLLR |
| | β-lactoglobulin | IDALNENK, TPEVDDEALEK, VYVEELKPTPEGDLEILLQK |
| Egg | Ovalbumin | GGLEPINFQTAADQAR, HIATNAVLFFGR |
| | Ovotransferrin | SAGWNIPIGTLIHR, FYTVISSLK |
| | Vitellogenin-1 | NVNFDGEILK, TVIVEAPIHGLK |
| Hazelnut | Cor a 9 | ADIYTEQVGR, ALPDDVLANAFQISR, LNALEPTNR, TNDNAQISPLAGR |
| Peanut | Ara h 1 | GSEEEDITNPINLR, GSEEEGDITNPINLR, GTGNLELVAVR |

The last step in the method development was the optimization of the collision energy (expressed in V) used to fragment each peptide in the collision cell of the triple quadrupole mass spectrometer. This collision energy was optimized to obtain the highest signal for the different transitions of each peptide and thus to obtain a higher method sensitivity. The mass spectrometers used at CER Groupe and ILVO are quite similar. They are both triple quadrupole mass spectrometers from Waters and have the same ionization source. However, the collision cells are different. This optimization had therefore to be performed in both laboratories.

A pseudo-optimal collision energy was proposed by the Skyline software, based on the peptide sequences. For each transition of the 19 peptide biomarkers, a range of collision energies going from -5 V to +5 V around the theoretically optimal value proposed by Skyline was evaluated to find the optimal collision energy. This window was extended if no optimal collision energy was identified. Example of collision energy optimization for peptide YLGYLEQLLR from milk $\alpha_{s1}$-casein is presented in Figure 62.

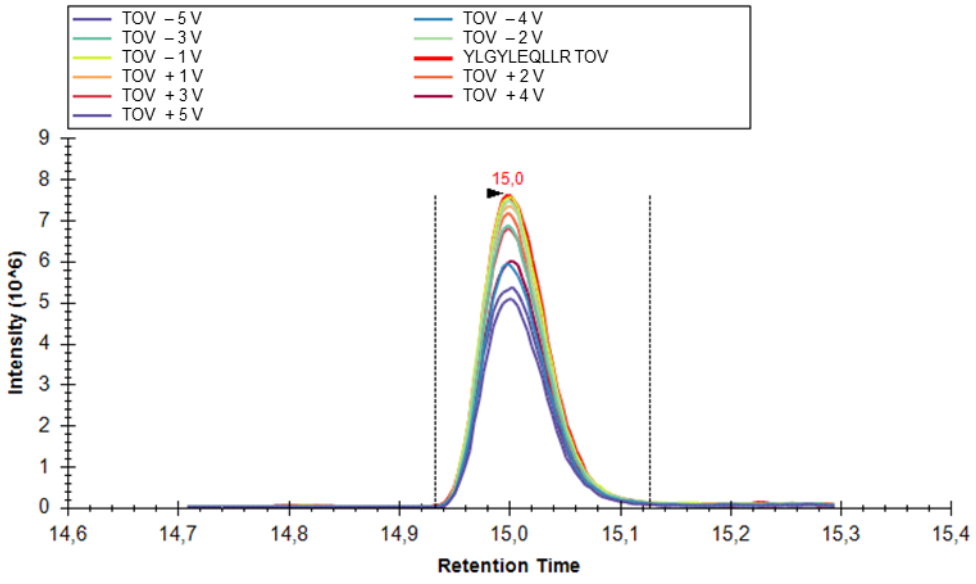



*Figure 62 - Identification of the optimal collision energy for peptide YLGYLEQLLR from milk $\alpha_{s1}$-casein. Collision energies ranging from -5 V to + 5 V around the theoretically optimal value (TOV) proposed by Skyline were investigated.*

The complete MRM method, with all 19 peptide biomarkers, the selected transitions for analyte and internal standard and the optimized collision energy (for the CER instrument) is presented in Table 13.

*Table 13 - Targeted peptides and associated multiple reaction monitoring (MRM) parameters for the UHPLC-MS/MS analysis. The most intense product ions, used for method validation calculations are highlighted in bold.*

| Allergen | Protein | Peptide | Precursor | Product ions (fragments) | Collision energy (V) |
|---|---|---|---|---|---|
| Milk | $\alpha_{s1}$-casein | FFVAPFPEVFGK | 693.3 (++) | 992.2 [y9], **921.1 [y8]**, 676.8 [y6] | 16 |
| | | (U-$^{15}$N) FFVAPFPEVFGK | 699.8 (++) | **930.0 [y8]** | |
| | | HQGLPQEVLNENLLR | 587.7 (+++) | 872.0 [y7], 790.8 [b7], **758.8 [y6]** | 12 |
| | | (U-$^{15}$N) HQGLPQEVLNENLLR | 595.6 (+++) | **769.8 [y6]** | |
| | | YLGYLEQLLR | 634.7 (++) | **992.2 [y8]**, 771.9 [y6], 658.8 [y5] | 16 |
| | | (U-$^{15}$N) YLGYLEQLLR | 641.7 (++) | **1004.1 [y8]** | |
| | β-lactoglobulin | IDALNENK | 459 (++) | 803.8 [y7], **688.8 [y6]**, 504.5 [y4] | 13 |
| | | (U-$^{15}$N) IDALNENK | 464.5 (++) | **697.7 [y6]** | |
| | | TPEVDDEALEK | 623.7 (++) | 919.0 [y8], 819.8 [y7], **573.1 [y10]** | 19 |
| | | (U-$^{15}$N) TPEVDDEALEK | 629.6 (++) | **578.6 [y10]** | |
| | | VYVEELKPTPEGDLEILLQK | 772.2 (+++) | **1026.7 [y18]**, 977.1 [y17], 628.2 [y11] | 17 |
| | | (U-$^{15}$N) VYVEELKPTPEGDLEILLQK | 779.8 (+++) | **1037.1 [y18]** | |
| Egg | Ovalbumin | GGLEPINFQTAADQAR | 844.9 (++) | 1122.2 [y10], 732.8 [y7], 666.7 [y12] | 22 |
| | | (U-$^{15}$N) GGLEPINFQTAADQAR | 855.8 (++) | **675.7 [y12]** | |
| | | HIATNAVLFFGR | 673.8 (++) | **1096.3 [y10]**, 1025.2 [y9], 924.1 [y8] | 22 |
| | | (U-$^{15}$N) HIATNAVLFFGR | 682.7 (++) | **1110.2 [y10]** | |
| | Ovotransferrin | SAGWNIPIGTLIHR | 512.6 (+++) | **907.1 [y8]**, 696.8 [y6], 538.7 [y4] | 15 |
| | | (U-$^{15}$N) SAGWNIPIGTLIHR | 519.6 (+++) | **920.0 [y8]** | |
| | | FYTVISSLK | 529.6 (++) | **747.9 [y7]**, 646.8 [y6], 547.7 [y5] | 12 |
| | | (U-$^{15}$N) FYTVISSLK | 534.6 (++) | **755.9 [y7]** | |
| | Vitellogenin-1 | NVNFDGEILK | 575.1 (++) | **936.0 [y8]**, 674.8 [y6], 559.7 [y5] | 13 |
| | | (U-$^{15}$N) NVNFDGEILK | 581.6 (++) | **946.0 [y8]** | |
| | | TVIVEAPIHGLK | 639.3 (++) | 865.0 [y8], 735.9 [y7], **664.8 [y6]** | 19 |
| | | (U-$^{15}$N) TVIVEAPIHGLK | 646.7 (++) | **673.8 [y6]** | |
| Hazelnut | Cor a 9 | ADIYTEQVGR | 576.6 (++) | 852.9 [y7], **689.7 [y6]**, 588.6 [y5] | 16 |
| | | (U-$^{15}$N) ADIYTEQVGR | 583.6 (++) | **699.7 [y6]** | |
| | | ALPDDVLANAFQISR | 815.9 (++) | 907.0 [y8], 835.9 [y7], **723.8 [y13]** | 19 |
| | | (U-$^{15}$N) ALPDDVLANAFQISR | 825.8 (++) | **732.7 [y13]** | |
| | | LNALEPTNR | 514.6 (++) | 729.8 [y6], 616.6 [y5], **487.5 [y4]** | 14 |
| | | (U-$^{15}$N) LNALEPTNR | 521.5 (++) | **495.5 [y4]** | |
| | | TNDNAQISPLAGR | 679.2 (++) | 713.8 [y7], **600.7 [y6]**, 513.6 [y5] | 19 |
| | | (U-$^{15}$N) TNDNAQISPLAGR | 688.7 (++) | **609.6 [y6]** | |
| Peanut | Ara h 1 | GSEEEDITNPINLR | 794.3 (++) | 828.0 [y7], 726.8 [y6], **612.7 [y5]** | 19 |
| | | (U-$^{15}$N) GSEEEDITNPINLR | 803.8 (++) | **621.7 [y5]** | |
| | | GSEEEGDITNPINLR | 822.9 (++) | 828.0 [y7], 726.8 [y6], **612.7 [y5]** | 22 |
| | | (U-$^{15}$N) GSEEEGDITNPINLR | 832.8 (++) | 8**621.7 [y5]** | |
| | | GTGNLELVAVR | 565.2 (++) | 800.0 [y7], 686.8 [y6], **557.7 [y5]** | 18 |
| | | (U-$^{15}$N) GTGNLELVAVR | 572.6 (++) | **565.7 [y5]** | |

## 5. Method validation (Results obtained by Maxime Gavage at CER Groupe)

Different performance parameters, including selectivity, LOD, LOQ, linearity, trueness and precision, were evaluated to validate the developed UHPLC-MS/MS method. In the absence of legislation, obtained results were compared to standard method requirements published by AOAC International in 2016 (Paez *et al*, 2016). This document describes the minimum recommended performance characteristics to be used for the detection and quantification by mass spectrometry of egg, milk, peanut, and hazelnut food allergens in finished food products and ingredients. These requirements are summarized in Table 14, already presented in the introduction chapter.

*Table 14 – Standard method performance requirements for detection and quantification of selected food allergens" (AOAC SMPR 2016.002)*

| | Target allergen | | | |
|---|---|---|---|---|
| **Parameter** | Whole egg | Milk | Peanut | Hazelnut |
| Analytical range (ppm) | 10-1000 | 10-1000 | 10-1000 | 10-1000 |
| MQL (ppm) | ≤ 5 | ≤ 10 | ≤ 10 | ≤ 10 |
| MDL (ppm) | ≤ 1.65 | ≤ 3 | ≤ 3 | ≤ 3 |
| Recovery (%) | 60 - 120 | 60 - 120 | 60 - 120 | 60 - 120 |
| $RSD_r$ (%) | ≤ 20 | ≤ 20 | ≤ 20 | ≤ 20 |
| $RSD_R$ (%) | ≤ 30 | ≤ 30 | ≤ 30 | ≤ 30 |

## 5.1. Data treatment

Acquired UHPLC-MS/MS data were processed using the TargetLynx software. The most intense transition (highlighted in bold in Table 13) was used for validation performance parameter calculations and the two others as confirmatory transitions. Calculations were based on chromatographic peak area ratio (for the most intense transition) between the peptide from the allergenic ingredient and its corresponding isotopically labelled version from the concatemer internal standard.

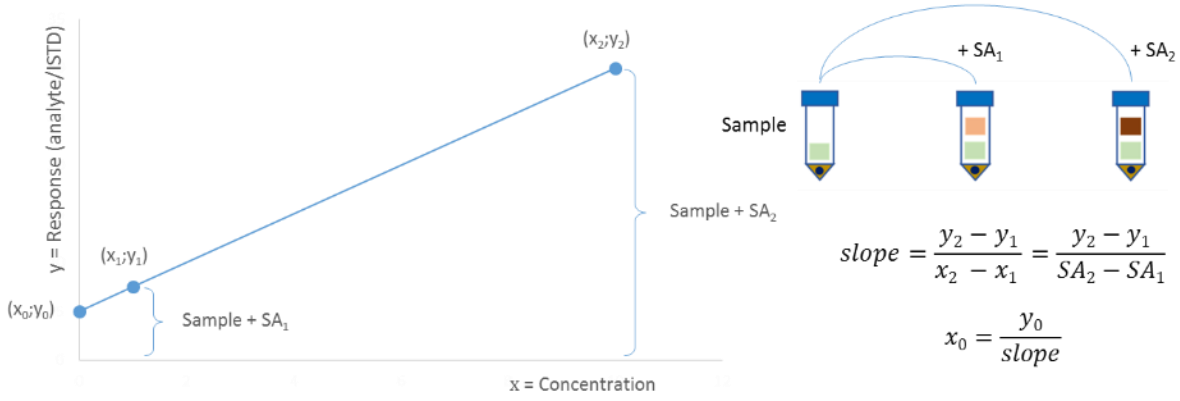



*Figure 63 – Standard addition quantification method. The allergen concentration of the sample $(x_0)$ is evaluated using the response obtained for this sample and the slope of the line formed by the two standard additions $(SA_1$ and $SA_2)$ performed on this sample.*

The standard addition quantification method was used for the evaluation of method trueness and precision. For a given sample, assuming method linearity, three aliquots were prepared and analysed. Defined and different standard amounts were added in two of them. As presented in Figure 63, allergen concentration is calculated based on the response obtained for this sample and the slope of the straight line obtained from the response of the two samples with standard addition.

## 5.2. Selectivity

Selectivity was already addressed during the method development (by Kaatje Van Vlierberghe at ILVO) with the analysis of blank and incurred test matrices. Final peptide biomarkers were selected based on sensitivity and selectivity criteria from the list of HRMS identified potential peptide biomarkers.

Selectivity was validated with the ability of the method to discriminate, for each peptide, the signal of the three selected transitions of the analyte and the corresponding one for the $^{15}N$ isotope-labelled internal standard from other components of blank matrices. The example of validated selectivity of the method for peptide GGLEPINFQTAADQAR of ovalbumin egg white protein in cookie is presented on Figure 64. No peak was observed at the retention time of the peptide for any of the four transitions.

Selectivity was evaluated for the selected transitions and the corresponding internal standard ones of the 19 peptide biomarkers in the three matrices (cookie, chocolate chips and unbaked freeze-dried cookie dough). Interference was observed for peptide HQGLPQEVLNENLLR of milk $\alpha_{s1}$-casein in chocolate chips (Figure 65) and peptide IDALNENK of milk β-lactoglobulin in unbaked freeze-dried cookie dough (Figure 66). These two combinations of interfering peptide and matrix were therefore excluded from the validation.

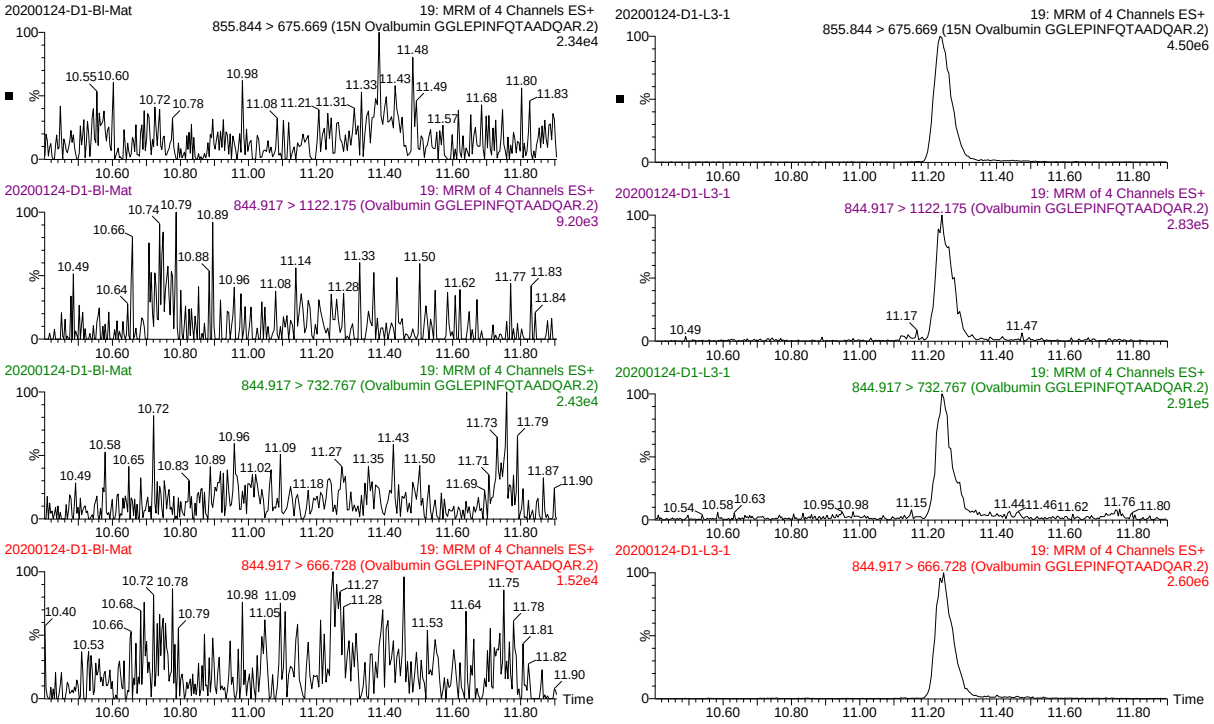



*Figure 64 – Validation of method selectivity, example for peptide GGLEPINFQTAADQAR of ovalbumin egg white protein in cookie. UHPLC-MS/MS analysis of blank cookie (on the left) and cookie spiked at 25 ppm with egg (on the right). Selectivity was validated for the transition (at the top) of the $^{15}N$ isotope-labelled internal standard and three transitions of the analyte.*

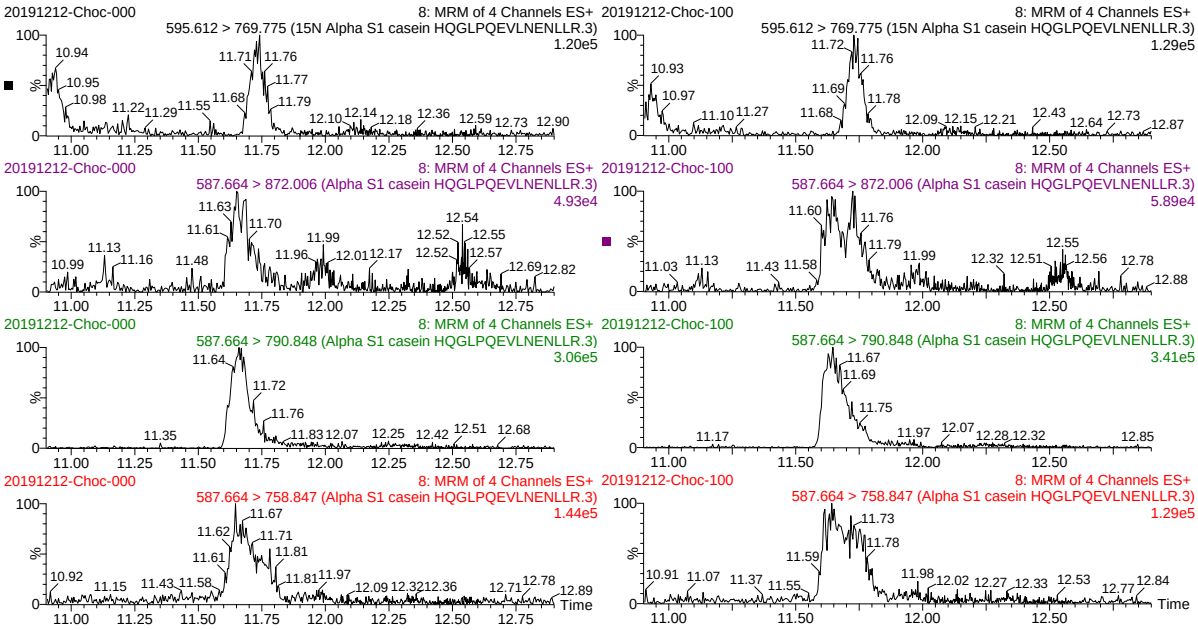



*Figure 65 – Interference observation for peptide HQGLPQEVLNENLLR of milk $\alpha_{s1}$-casein in chocolate chips matrix. UHPLC-MS/MS analysis of blank matrix (on the left) and matrix spiked at 10 ppm with milk (on the right). $^{15}$N isotope-labelled concatemer internal standard (transitions at the top with black label) was spiked in both samples.*

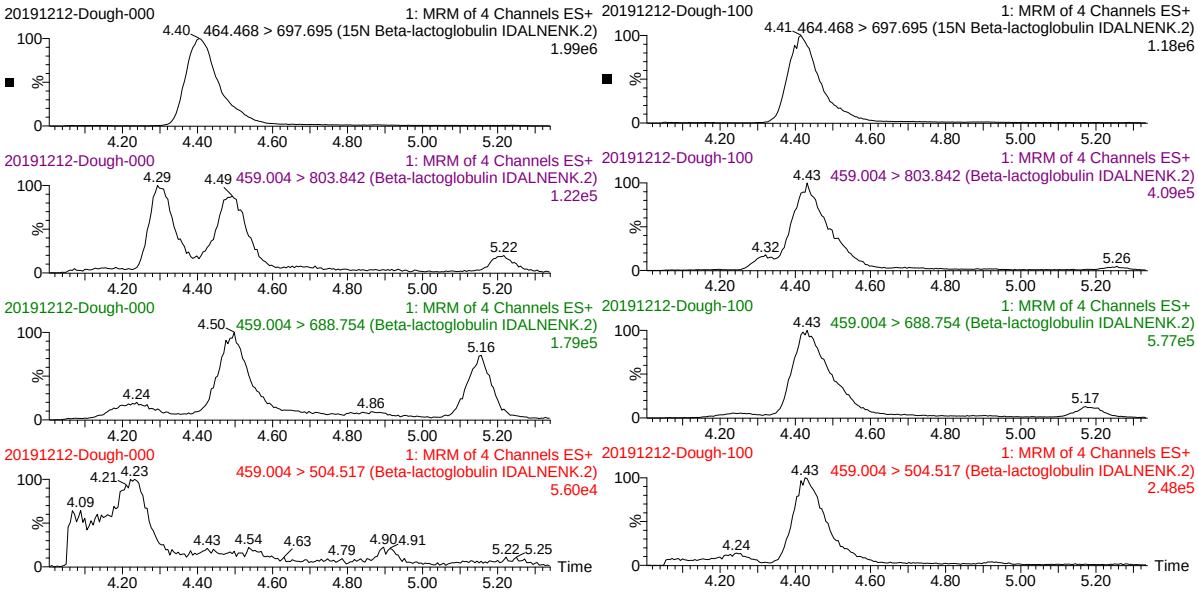



*Figure 66 - Interference observation for peptide IDALNENK of milk ß-lactoglobulin in unbaked freeze-dried cookie dough matrix. UHPLC-MS/MS analysis of blank matrix (on the left) and matrix spiked at 10 ppm with milk (on the right). $^{15}$N isotope-labelled concatemer internal standard (transitions at the top with black label) was spiked in both samples.*

## 5.3. LOD and LOQ

LOD and LOQ of the method used were evaluated for each one of the three matrices incurred with the four allergenic ingredients at different contamination levels (0.5, 1, 2.5, 5, 10 and 25 ppm). Incurred matrices were used to integrate the impact of food processing on allergen detectability. Each contamination was analysed with three independent replicates. The LOD of the method for a given peptide corresponded to the lowest contamination level for which S/N > 3 for the three replicates. The same procedure was applied for LOQ with S/N > 10.

Example of S/N calculation for peptide ADIYTEQVGR from hazelnut in cookie incurred at the 5 ppm level is exposed on Figure 67.

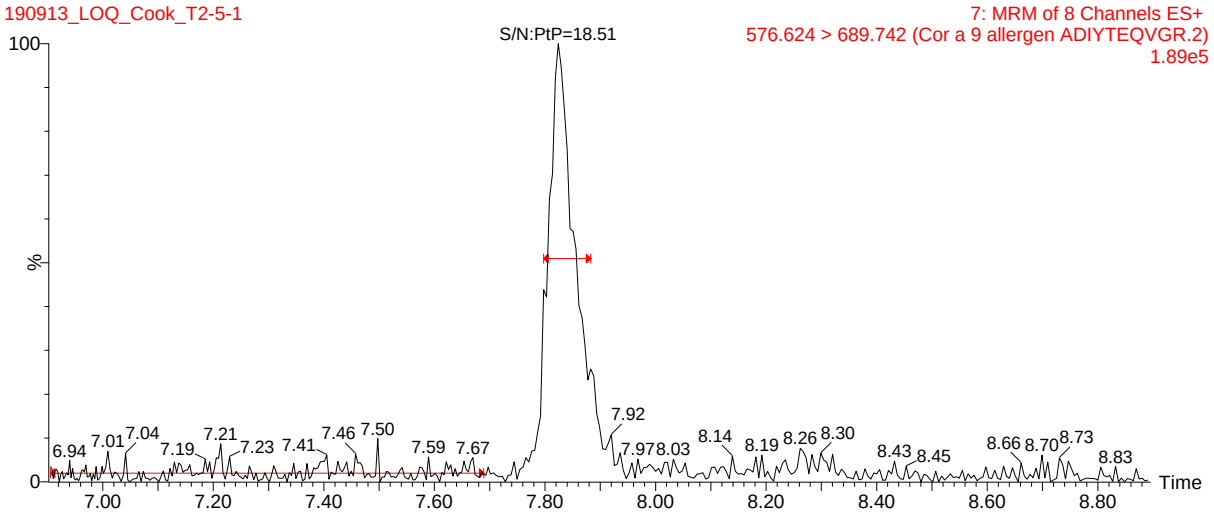



*Figure 67 – Example of S/N calculation using the peak-to-peak method for peptide ADIYTEQVGR from hazelnut in cookie incurred at the 5 ppm level.*

Complete results of LOD and LOQ determination are presented in Table 15.

*Table 15 – LOD and LOQ determination for each one of the 19 peptide biomarkers based on the UHPLC-MS/MS analysis of three independent replicates for each matrix. LOD of the method for a given peptide corresponded to the lowest contamination level for which S/N > 3 for the three replicates. The same procedure was applied for LOQ with S/N > 10.*

| Allergen | Protein | Peptide | Cookie | | Chocolate | | Dough | |
|---|---|---|---|---|---|---|---|---|
| | | | LOD | LOQ | LOD | LOQ | LOD | LOQ |
| Milk | $\alpha_{s1}$-casein | FFVAPFPEVFGK | 0.5 ppm | 0.5 ppm | 0.5 ppm | 0.5 ppm | 0.5 ppm | 0.5 ppm |
| | | HQGLPQEVLNENLLR | 0.5 ppm | 2.5 ppm | / | / | 0.5 ppm | 2.5 ppm |
| | | YLGYLEQLLR | 0.5 ppm | 0.5 ppm | 0.5 ppm | 0.5 ppm | 0.5 ppm | 1 ppm |
| | β-lactoglobulin | IDALNENK | 0.5 ppm | 5 ppm | 0.5 ppm | 0.5 ppm | / | / |
| | | TPEVDDEALEK | 2.5 ppm | 25 ppm | 5 ppm | 10 ppm | 0.5 ppm | 2.5 ppm |
| | | VYVEELKPTPEGDLEILLQK | 2.5 ppm | 25 ppm | 0.5 ppm | 0.5 ppm | 1 ppm | 2.5 ppm |
| Egg | Ovalbumin | GGLEPINFQTAADQAR | 1 ppm | 10 ppm | 25 ppm | > 25 ppm | 2.5 ppm | 10 ppm |
| | | HIATNAVLFFGR | 25 ppm | > 25 ppm | > 25 ppm | > 25 ppm | 10 ppm | > 25 ppm |
| | Ovotransferrin | SAGWNIPIGTLIHR | 10 ppm | 25 ppm | 5 ppm | 10 ppm | 10 ppm | > 25 ppm |
| | | FYTVISSLK | > 25 ppm | > 25 ppm | > 25 ppm | > 25 ppm | > 25 ppm | > 25 ppm |
| | Vitellogenin-1 | NVNFDGEILK | > 25 ppm | > 25 ppm | 10 ppm | > 25 ppm | 25 ppm | > 25 ppm |
| | | TVIVEAPIHGLK | > 25 ppm | > 25 ppm | > 25 ppm | > 25 ppm | > 25 ppm | > 25 ppm |
| Hazelnut | Cor a 9 | ADIYTEQVGR | 0.5 ppm | 2.5 ppm | 10 ppm | > 25 ppm | 1 ppm | 5 ppm |
| | | ALPDDVLANAFQISR | 0.5 ppm | 1 ppm | 2.5 ppm | 25 ppm | 0.5 ppm | 1 ppm |
| | | LNALEPTNR | 0.5 ppm | 2.5 ppm | 25 ppm | > 25 ppm | 5 ppm | 25 ppm |
| | | TNDNAQISPLAGR | 0.5 ppm | 2.5 ppm | 10 ppm | > 25 ppm | 0.5 ppm | 5 ppm |
| Peanut | Ara h 1 | GSEEEDITNPINLR | 25 ppm | > 25 ppm | 25 ppm | > 25 ppm | 25 ppm | > 25 ppm |
| | | GSEEEGDITNPINLR | 25 ppm | > 25 ppm | 25 ppm | > 25 ppm | 25 ppm | > 25 ppm |
| | | GTGNLELVAVR | 1 ppm | 5 ppm | 2.5 ppm | 25 ppm | 1 ppm | 5 ppm |

Given the observed interference for peptides HQGLPQEVLNENLLR and IDALNENK from milk in chocolate chips and freeze-dried cookie dough matrices respectively, no result was presented for these two peptides.

Obtained results were summarized in Table 16, considering peptides with the lowest LOD and LOQ for each allergenic ingredient, and compared to AOAC requirements. The developed UHPLC-MS/MS method was able to detect and quantify milk in all three matrices with AOAC sensitivity requirements. On the contrary, egg detection was globally not sensitive enough. Difficulties were also observed for the chocolate chips matrix, which is known to impact sample preparation (Korte *et al*, 2019).

*Table 16 - Comparison of obtained LOD and LOQ for each allergenic ingredient and comparison with AOAC requirements (green in agreement with AOAC requirements - red not in agreement with AOAC requirements)*

| | Cookie | | Chocolate | | Dough | |
|---|---|---|---|---|---|---|
| | LOD | LOQ | LOD | LOQ | LOD | LOQ |
| Milk | 0.5 ppm | 0.5 ppm | 0.5 ppm | 0.5 ppm | 0.5 ppm | 1 ppm |
| Egg | 1 ppm | 10 ppm | 1 ppm | 10 ppm | 2.5 ppm | 10 ppm |
| Peanut | 1 ppm | 5 ppm | 2.5 ppm | 25 ppm | 1 ppm | 5 ppm |
| Hazelnut | 0.5 ppm | 1 ppm | 2.5 ppm | 25 ppm | 0.5 ppm | 1 ppm |

### 5.4. Linearity

To evaluate method linearity, blank matrices were spiked with defined levels of allergen extracts. This strategy was chosen to eliminate the variability due to allergen content in incurred matrices. Standard reference material, with characterized protein content were used. Spray-dried whole egg (RM 8445 from NIST), skim milk powder (SPM-MQA-092104 from MoniQA), light roasted and partially defatted peanut flour (LGCQ1020 from LGC Standard) and unroasted hazelnut floor (from iFAAM project) were considered. Allergen extracts were separately obtained with the same extraction protocol than for the sample preparation described above. Knowing the protein content of the different standards and assuming a 100 % protein extraction yield, standard and extraction buffer ratio was adapted to obtain a 20 mg/ml protein concentration. Standard extracts were further diluted and mixed to spike 2 g sample with a 100 µl volume.

Linearity was evaluated for each one of the 19 peptide biomarkers. For each one of the three food matrices (cookie, chocolate chips and freeze-dried cookie dough), a single calibration curve including 0, 0.5, 1, 2.5, 5, 10, 25 and 50 ppm contamination level was prepared and analysed. Linearity assessment was based the calculation of the coefficient of determination ($R^2$) using linear regression with 1/X calibration weighing function.

Example of linearity evaluation for peptide FFVAPFPEVFGK from milk is exposed on Figure 68.

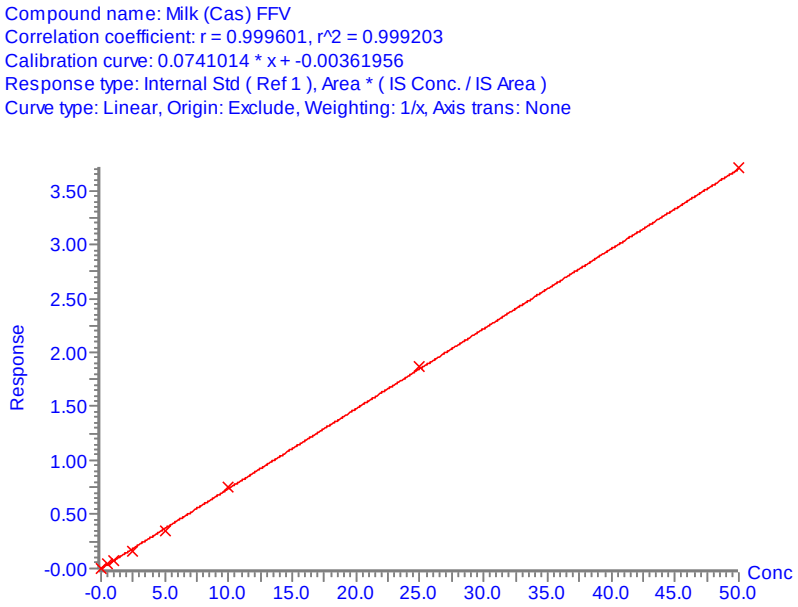



*Figure 68 – Linearity evaluation of peptide FFVAPFPEVFGK from milk using 1/X weighing function.*

Complete results are exposed in Table 17. It could be assumed that linearity was obtained in the analysed contamination range. Linearity was evaluated by visual inspection of a plot of response as a function of analyte concentration and the majority, *i.e.* 40/45 (certain combinations of peptides and matrices were excluded due to interference or insufficient sensitivity) of obtained $R^2$ were higher than 0.99.

With the use of spiked samples, allergenic ingredients were not subject to food processing. Observed LOD (S/N > 3) was generally lower compared to incurred samples.

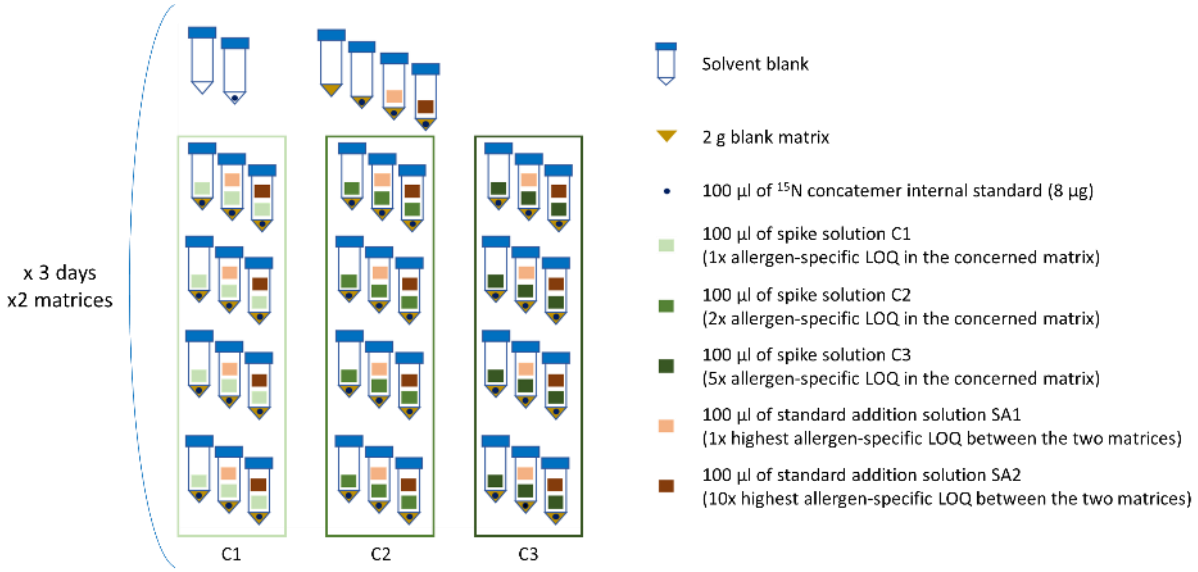
Interference was again observed for peptides HQGLPQEVLNENLLR and IDALNENK from milk in chocolate chips and freeze-dried cookie dough matrices respectively. Linearity was therefore not evaluated for these two peptides.

*Table 17 – Evaluation of UHPLC-MS/MS linearity with the analysis of three food matrices (cookie, chocolate chips and freeze-dried cookie dough) spiked at different levels (from 0.5 to 50 ppm) with allergenic ingredient standard extracts. Coefficient of determination ($R^2$) using linear regression with 1/X calibration weighing function was calculated for the points of the calibration curve with S/N > 3.*

| Allergen | Protein | Peptide | Cookie | | | Chocolate | | | Dough | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Lowest calibration level (S/N > 3) | Number of calibration levels | $R^2$ | Lowest calibration level (S/N > 3) | Number of calibration levels | $R^2$ | Lowest calibration level (S/N > 3) | Number of calibration levels | $R^2$ |
| Milk | $\alpha_{s1}$-casein | FFVAPFPEVFGK | 0.5 ppm | 8 | 0.9985 | 0.5 ppm | 8 | 0.9992 | 0.5 ppm | 8 | 0.9979 |
| | | HQGLPQEVLNENLLR | 0.5 ppm | 8 | 0.9898 | / | / | / | 1 ppm | 7 | 0.9935 |
| | | YLGYLEQLLR | 0.5 ppm | 8 | 0.9840 | 0.5 ppm | 8 | 0.9984 | 0.5 ppm | 8 | 0.9995 |
| | β-lactoglobulin | IDALNENK | 0.5 ppm | 8 | 0.9984 | 0.5 ppm | 8 | 0.9982 | / | / | / |
| | | TPEVDDEALEK | 2.5 ppm | 6 | 0.9963 | 2.5 ppm | 6 | 0.9994 | 0.5 ppm | 8 | 0.9977 |
| | | VYVEELKPTPEGDLEILLQK | 0.5 ppm | 8 | 0.9985 | 0.5 ppm | 8 | 0.9968 | 0.5 ppm | 8 | 0.9965 |
| Egg | Ovalbumin | GGLEPINFQTAADQAR | 2.5 ppm | 6 | 0.9957 | 10 ppm | 4 | 0.9848 | 2.5 ppm | 6 | 0.9840 |
| | | HIATNAVLFFGR | 10 ppm | 4 | 0.9942 | 50 ppm | 2 | / | 50 ppm | 2 | / |
| | Ovotransferrin | SAGWNIPIGTLIHR | 2.5 ppm | 6 | 0.9919 | 5 ppm | 5 | 0.9995 | 2.5 ppm | 6 | 0.9991 |
| | | FYTVISSLK | 5 ppm | 5 | 0.9966 | 50 ppm | 2 | / | 50 ppm | 2 | / |
| | Vitellogenin-1 | NVNFDGEILK | 2.5 ppm | 6 | 0.9996 | 2.5 ppm | 6 | 0.9992 | 10 ppm | 4 | 0.9999 |
| | | TVIVEAPIHGLK | 10 ppm | 4 | 0.9975 | 5 ppm | 5 | 0.9961 | 50 ppm | 2 | / |
| Hazelnut | Cor a 9 | ADIYTEQVGR | 0.5 ppm | 8 | 0.9980 | 0.5 ppm | 8 | 0.9981 | 0.5 ppm | 8 | 0.9986 |
| | | ALPDDVLANAFQISR | 0.5 ppm | 8 | 0.9977 | 0.5 ppm | 8 | 0.9989 | 0.5 ppm | 8 | 0.9994 |
| | | LNALEPTNR | 0.5 ppm | 8 | 0.9990 | 2.5 ppm | 6 | 0.9988 | 2.5 ppm | 6 | 0.9928 |
| | | TNDNAQISPLAGR | 0.5 ppm | 8 | 0.9973 | 0.5 ppm | 8 | 0.9986 | 2.5 ppm | 6 | 0.9950 |
| Peanut | Ara h 1 | GSEEEDITNPINLR | 50 ppm | 2 | / | 50 ppm | 2 | / | 50 ppm | 2 | / |
| | | GSEEEGDITNPINLR | 50 ppm | 2 | / | 50 ppm | 2 | / | 50 ppm | 2 | / |
| | | GTGNLELVAVR | 5 ppm | 5 | 0.9931 | 10 ppm | 4 | 0.9951 | 5 ppm | 5 | 0.9770 |

### 5.5. Trueness and precision

As for linearity, trueness and precision were evaluated using blank matrices spiked with allergen extracts to eliminate the variability due to allergen content in incurred matrices. Trueness and precision were evaluated in cookie and chocolate matrices using the standard addition quantification method. The analysis of three contamination levels, corresponding to 1x, 2x and 5x allergen-specific LOQ (with spiked allergen extracts) in the concerned matrix, were considered. Standard addition was performed by spiking samples with 1x and 10x the highest allergen-specific LOQ (with spiked allergen extracts). Four independent replicates were prepared and analysed for each contamination level. The whole procedure was repeated for three consecutive days. Each day, solvent blanks (with and without internal standard) and matrix blanks (with and without internal standard) were included. Contamination levels, standard addition levels for each allergen and illustration of samples prepared for one day of trueness and precision evaluation is exposed on Figure 69.




| | Validation concentrations | | | | | | Standard addition levels | |
|---|---|---|---|---|---|---|---|---|
| | Cookie | | | Chocolate | | | | |
| | C1 | C2 | C3 | C1 | C2 | C3 | SA1 | SA2 |
| Egg | 5 ppm | 10 ppm | 25 ppm | 10 ppm | 20 ppm | 50 ppm | 10 ppm | 100 ppm |
| Milk | 0.5 ppm | 1 ppm | 2.5 ppm | 0.5 ppm | 1 ppm | 2.5 ppm | 0.5 ppm | 5 ppm |
| Peanut | 10 ppm | 20 ppm | 50 ppm | 25 ppm | 50 ppm | 125 ppm | 25 ppm | 250 ppm |
| Hazelnut | 1 ppm | 2 ppm | 5 ppm | 1 ppm | 2 ppm | 5 ppm | 1 ppm | 10 ppm |

*Figure 69 – Schematic overview of samples prepared for trueness and precision evaluation using standard addition quantification method. Spike levels (C1, C2 and C3) and standard addition levels (SA1 and SA2) are detailed for each allergenic ingredient and food matrix.*

For each peptide biomarker, each matrix and each of the 3 validation levels, trueness and precision were evaluated. Method trueness evaluation was based on measured recovery considering theoretical contamination level. Precision was evaluated on repeatability ($RSD_r$) and a parameter associated to intermediate precision ($RSD_{INT}$) even if the analyst was identical. This $RSD_{INT}$ could be seen as a "within-lab reproducibility" (González *et al*, 2010).

The precision of a method can indeed be evaluated with three parameters corresponding to relative standard deviation under different conditions exposed in Table 18.

*Table 18 – Different levels of method precision evaluation*

| | **Repeatability** | **Intermediate precision** | **Reproducibility** |
|---|---|---|---|
| | $RSD_r$ | $RSD_{INT}$ | $RSD_R$ |
| | n = 3-5 on single day | n = 3-5 on 2 or 3 days | |
| **Lab location** | Identical | Identical | Different |
| **Sample** | Identical | Identical | Identical |
| **Analyst** | Identical | Different | Different |
| **Instrument** | Identical | Identical/different | Different |
| **Time period** | Short | Medium | Long |
| **Standards** | Identical | Different | Different |
| **Reagents** | Identical | Different | Different |

Complete results of trueness and precision evaluation were presented in Table 19 and Table 20 for cookie and chocolate chips matrices respectively. Contamination levels were based on spiked extracts LOQ for the different allergenic ingredients. All peptides were therefore not identified at such low contamination levels and no result concerning trueness and precision were obtained.

Standard method performance requirements from AOAC International were globally achieved with a couple of exceptions, mainly at the lower contamination level (C1) and for chocolate chips matrix. However, for each allergenic ingredient and the two considered matrices, at least one peptide fulfils AOAC requirements regarding trueness and precision. These peptides should be used for the quantification.

*Table 19 – Evaluation of UHPLC-MS/MS precision ($RSD_r$ and $RSD_{INT}$) and trueness (recovery) in cookie matrix using the standard addition method. Cookie 2 g samples were spiked with allergen extracts at 3 different levels (C1, C2 and C3 corresponding to 1x, 2x and 5x allergen specific LOQ). For each contamination level, four biological replicates were prepared and analysed. The whole procedure was repeated for three consecutive days. Trueness and precision results were compared with AOAC requirements (green in agreement with AOAC requirements red not in agreement with AOAC requirements).*

| Allergen | Protein | Peptide | C1 (1x LOQ) | | | C2 (2x LOQ) | | | C3 (5x LOQ) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $RSD_r$ (%) | $RSD_{INT}$ (%) | Recovery (%) | $RSD_r$ (%) | $RSD_{INT}$ (%) | Recovery (%) | $RSD_r$ (%) | $RSD_{INT}$ (%) | Recovery (%) |
| Milk | $\alpha_{s1}$-casein | FFVAPFPEVFGK | 15.4 | 18.4 | 107.1 | 2.7 | 3.9 | 101.1 | 5.2 | 5.0 | 105.8 |
| | | HQGLPQEVLNENLLR | 10.7 | 13.3 | 132.5 | 8.4 | 8.8 | 109.9 | 16.2 | 16.5 | 108.0 |
| | | YLGYLEQLLR | 13.7 | 14.4 | 109.8 | 3.1 | 3.5 | 102.2 | 3.1 | 3.5 | 102.2 |
| | β-lactoglobulin | IDALNENK | 5.4 | 6.9 | 128.9 | 2.5 | 6.9 | 108.6 | 3.5 | 7.5 | 95.2 |
| | | TPEVDDEALEK | / | / | / | / | / | / | / | / | / |
| | | VYVEELKPTPEGDLEILLQK | 6.3 | 6.8 | 117.4 | 4.8 | 7.0 | 102.6 | 7.3 | 8.5 | 97.5 |
| Egg | Ovalbumin | GGLEPINFQTAADQAR | 3.7 | 10.1 | 80.4 | 4.2 | 12.1 | 85.2 | 5.4 | 16.0 | 85.2 |
| | | HIATNAVLFFGR | 6.1 | 11.2 | 83.3 | 7.2 | 16.3 | 79.3 | 9.3 | 22.0 | 74.4 |
| | Ovotransferrin | SAGWNIPIGTLIHR | 5.1 | 11.1 | 74.4 | 4.9 | 12.8 | 83.1 | 11.3 | 20.5 | 86.4 |
| | | FYTVISSLK | 9.3 | 8.6 | 128.4 | 9.8 | 11.7 | 122.4 | 7.5 | 13.9 | 103.7 |
| | Vitellogenin-1 | NVNFDGEILK | 6.9 | 6.5 | 101.6 | 7.4 | 7.2 | 102.9 | 2.7 | 3.9 | 98.1 |
| | | TVIVEAPIHGLK | / | / | / | / | / | / | / | / | / |
| Hazelnut | Cor a 9 | ADIYTEQVGR | 4.7 | 8.9 | 99.2 | 2.3 | 7.8 | 89.8 | 3.8 | 8.1 | 85.1 |
| | | ALPDDVLANAFQISR | 5.4 | 6.0 | 103.8 | 3.3 | 8.6 | 95.0 | 4.5 | 9.3 | 89.2 |
| | | LNALEPTNR | 7.0 | 10.8 | 118.6 | 5.5 | 11.8 | 101.9 | 7.8 | 10.8 | 94.8 |
| | | TNDNAQISPLAGR | 11.3 | 16.4 | 150.1 | 7.7 | 17.5 | 115.2 | 6.8 | 10.1 | 94.0 |
| Peanut | Ara h 1 | GSEEEDITNPINLR | / | / | / | / | / | / | / | / | / |
| | | GSEEEGDITNPINLR | / | / | / | / | / | / | / | / | / |
| | | GTGNLELVAVR | 6.7 | 8.2 | 121.2 | 6.9 | 8.0 | 106.7 | 4.6 | 6.7 | 91.7 |

*Table 20 – Evaluation of UHPLC-MS/MS precision ($RSD_r$ and $RSD_{INT}$) and trueness (recovery) in chocolate chip matrix using the standard addition method. Chocolate chip 2 g samples were spiked with allergen extracts at 3 different levels (C1, C2 and C3 corresponding to 1x, 2x and 5x allergen specific LOQ). For each contamination level, four biological replicates were prepared and analysed. The whole procedure was repeated for three consecutive days. Trueness and precision results were compared with AOAC requirements (green in agreement with AOAC requirements red not in agreement with AOAC requirements).*

| Allergen | Protein | Peptide | C1 (1x LOQ) | | | C2 (2x LOQ) | | | C3 (5x LOQ) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $RSD_r$ (%) | $RSD_{INT}$ (%) | Recovery (%) | $RSD_r$ (%) | $RSD_{INT}$ (%) | Recovery (%) | $RSD_r$ (%) | $RSD_{INT}$ (%) | Recovery (%) |
| Milk | $\alpha_{s1}$-casein | FFVAPFPEVFGK | 7.3 | 16.2 | 116.6 | 12.3 | 11.9 | 113.5 | 12.8 | 11.3 | 102.0 |
| | | HQGLPQEVLNENLLR | / | / | / | / | / | / | / | / | / |
| | | YLGYLEQLLR | 24.4 | 35.3 | 137.1 | 17.5 | 17.4 | 114.8 | 14.6 | 15.1 | 107.0 |
| | β-lactoglobulin | IDALNENK | 4.5 | 8.8 | 158.4 | 14.7 | 13.5 | 131.3 | 5.2 | 5.2 | 107.0 |
| | | TPEVDDEALEK | / | / | / | / | / | / | / | / | / |
| | | VYVEELKPTPEGDLEILLQK | 10.8 | 13.9 | 144.3 | 14.2 | 14.9 | 123.8 | 7.8 | 10.5 | 112.2 |
| Egg | Ovalbumin | GGLEPINFQTAADQAR | 13.4 | 15.0 | 73.0 | 12.7 | 13.0 | 78.8 | 9.7 | 25.1 | 79.4 |
| | | HIATNAVLFFGR | 16.7 | 16.9 | 90.9 | 29.8 | 26.3 | 88.4 | 18.9 | 21.8 | 92.1 |
| | Ovotransferrin | SAGWNIPIGTLIHR | 10.6 | 62.6 | 159.7 | 7.3 | 65.1 | 150.4 | 10.6 | 54.3 | 170.9 |
| | | FYTVISSLK | / | / | / | / | / | / | / | / | / |
| | Vitellogenin-1 | NVNFDGEILK | 7.7 | 8.4 | 103.7 | 16.8 | 15.0 | 106.1 | 15.5 | 17.9 | 99.9 |
| | | TVIVEAPIHGLK | 20.3 | 18.5 | 113.4 | 15.2 | 13.9 | 121.1 | 23.2 | 23.4 | 120.4 |
| Hazelnut | Cor a 9 | ADIYTEQVGR | 15.3 | 17.9 | 113.2 | 10.9 | 14.5 | 97.8 | 21.7 | 19.8 | 110.2 |
| | | ALPDDVLANAFQISR | 6.7 | 6.0 | 101.7 | 4.3 | 7.3 | 96.4 | 3.3 | 7.9 | 98.1 |
| | | LNALEPTNR | / | / | / | / | / | / | / | / | / |
| | | TNDNAQISPLAGR | 11.2 | 14.1 | 134.9 | 7.6 | 7.6 | 124.0 | 12.5 | 11.3 | 104.7 |
| Peanut | Ara h 1 | GSEEEDITNPINLR | / | / | / | / | / | / | / | / | / |
| | | GSEEEGDITNPINLR | / | / | / | / | / | / | / | / | / |
| | | GTGNLELVAVR | 7.4 | 8.5 | 103.9 | 7.4 | 6.9 | 102.0 | 12.0 | 12.5 | 102.9 |

## 6. Conclusion

Based on potential peptide biomarkers identified by HRMS, a UHPLC-MS/MS method was developed. A total of 19 peptide biomarkers were selected in the list of potential biomarkers based on selectivity and sensitivity criteria to quantify the four considered allergenic ingredients (egg, milk, peanut and hazelnut) in processed food products.

The validation was the last step in the UHPLC-MS/MS method development. Combining the use of a $^{15}N$ isotope-labelled concatemer internal standard, incurred and spiked food matrices and standard addition quantification, performance parameters including selectivity, LOD, LOQ, linearity, trueness and precision were evaluated.

The obtained results were compared to standard method performances required by the AOAC International for the detection and the quantification of selected food allergens.

Method sensitivity was allergen and matrix dependent. Milk was detected with sufficient sensitivity whereas egg LOQ was higher than the AOAC requirement. As expected from known effects on sample preparation and analysis, chocolate chips matrices were also found to be problematic for egg, peanut and hazelnut detection with obtained LOQ higher than the AOAC requirement.

Method linearity was achieved in the 0.5 – 50 ppm contamination range with visually inspected linear calibration curves and $R^2$, coefficient of determination, higher than 0.99 for most of the peptides.

Method trueness and precision were finally evaluated with spiked matrices and standard addition quantification. Trueness was evaluated using recovery considering theoretical spiking values and precision with repeatability and a relative standard deviation close to intermediate precision definition. At least one peptide per matrix and per allergenic ingredient is in agreement with AOAC requirements.

**CONCLUSION AND PERSPECTIVES**

The prevalence of food allergy has increased during the last decades and is now recognized as a significant public health issue in developed countries (Savage & Johns, 2015). In the absence of validated treatment, the only solution for allergic patients is to strictly avoid to consume the culprit food. However, accidental allergen intake remains possible and symptoms associated to the allergic reaction can be life-threatening. Beside the risk of severe allergic reactions, food hypersensitivity has an impact on psychological distress and on the quality of life of allergic patients, as well as their families.

To protect allergic consumers, the legislation of numerous developed countries requires the labelling of several allergenic ingredients. At the European level, the labelling of a list of 14 allergenic ingredients and substances causing intolerances is required. The current legislation however only covers allergenic food when used as an ingredient, that is, foods or substances that are voluntarily incorporated into the food product, that are part of the recipe. This regulation does not address a major problem in food allergy management: cross-contamination.

This task is highly complex and zero risk in food allergy management cannot be achieved (Remington *et al*, 2020). Zero risk is not operationally achievable in the food industry given the complexities in the supply chain and in processing facilities that handle multiple allergenic foods, often on shared processing equipment. Good manufacturing practices and allergen management are established by the food industry (Gupta *et al*, 2017), however, ensuring that 100 % of all allergen residue is removed from shared equipment is impossible. Moreover, zero risk is also not achievable from an analytical point of view. Despite the improvement of detection methods sensitivity, presented detection limits will always be defined allergen values.

Since zero risk is not feasible from a risk management perspective for allergic individuals or the food industry, consensus on a worldwide accepted level of risk is needed. Thus, when possible, the aim should be for the protection of a certain proportion of the population, e.g. to protect 99% of the allergic population from any objective reaction (Remington *et al*, 2020). These Reference Doses (mg protein) were meant to serve as a basis for calculating action levels (concentrations, mg protein/kg) for precautionary allergen labelling. This approach afforded a possibly acceptable public health outcome by striking a balance between the statistical certainty of the estimate, level of consumer protection, ability to analytically verify and enforce, and hopeful enhancement of consumer observance and avoidance of food products with precautionary labelling. The experts from VITAL are developing such an approach since 2006.

Even if strongly requested by food producers and control laboratories, no harmonized regulatory framework for managing hidden allergens or action thresholds have been enacted in Europe. These aspects are essential for the development of quantitative risk assessment in allergen management. Different countries, including Belgium, proposed allergen reference doses to manage the risk that may arise from the presence of allergens in foods. The problem is that these thresholds differ greatly from country to country, even in Europe. The reference doses proposed by the NVWA in the Netherlands are for instance 10 to 100 times (depending on the allergen) lower than doses proposed in VITAL 3.0 and Belgium. These doses were proposed to protect even the most sensitive patients but are manageable with the current analytical methods. A global harmonized regulatory framework for managing hidden allergens is urgently needed.

Quantitative allergen analysis methods are essential for efficient allergen management. Allergen detection is currently performed with immunological assays that base their detection on antibody-allergen (or other marker protein) interaction and with PCR methods that target the DNA from

allergenic ingredients (Costa *et al*, 2016a). As presented in a critical review from Senyuva and co-workers (Senyuva *et al*, 2019), ELISA and PCR test kits were developed and are commercially available for the detection of the 13 allergenic ingredients that require labelling in Europe (sulphur dioxide is excluded from this list). The only exception is the analysis of celery with immunological methods due to cross-reactivity with proteins from other plants such as parsley, carrot, coriander or fennel (Flanagan, 2014). The claimed limits of detection of the different kits are in the range of the ppm and are in agreement with clinical thresholds. For some of the test kits, very specific individual foods are indicated for which the test kit has been validated, whereas in other cases very general indications are given such as applicable to "raw materials, uncooked/cooked foods". However, validation is crucial in the development of analysis methods. The use of incurred materials is absolutely required for the development and the validation of food allergen detection methods. Their use permits to reproduce conventional formulation, realistic incorporation of allergens inside the food microstructure and allergen modifications due to food processing, thus allowing more rigorous assessment of quantitative method performance in terms of extraction yield, accuracy, and sensitivity when processed foods are analysed (Mattarozzi & Careri, 2019).

Furthermore, immunological assays and PCR methods suffer from certain limitations. Besides matrix effect, performance of ELISA tests can be affected by food processing, which is predominantly attributed to tri-dimensional or chemical modifications of the target proteins (Costa *et al*, 2016a). In a review article from Iqbal and co-workers dedicated to peanut protein detection by ELISA (Iqbal *et al*, 2016), the authors highlighted that the detectability of peanut proteins was reduced due to heat treatment. The detectability was reduced by more than 80 % in the case of peanuts processed for a 12 min heat treatment at 100 °C and even by more than 90 % in the presence of 20 mM glucose (probably due to Maillard reactions). Immunological assays could also be subject to cross-reactivity. In a recent study, Koeberl and co-workers (Koeberl *et al*, 2018) compared three commercially available ELISA test kits for the detection of lupine. The three kits were able to detect three species of lupine, though with different sensitivities. However, all showed some cross-reactivity to related vegetable samples analysed such as peanut or soya.

Another limitation of the ELISA method is the impossible multiplexed allergen analysis. To overcome this limitation, innovative antibody-based solutions were developed such as the multi-analyte profiling (xMAP®) technology (Houser, 2012). Multiplexed analysis is performed thanks to the use of differentially detectable bead sets as a substrate capturing analytes in solution and employs fluorescent methods for detection. A single laboratory validation of the multiplex xMAP® food allergens (crustaceans, soy, nine tree nuts, egg, gluten, peanut and milk) detection with incurred food samples was performed by Nowatzke and co-workers (Nowatzke *et al*, 2019). Replicate samples displayed good reproducibility (RSD typically < 5%). However, the same general limitations of immunological methods concerning food processing and matrix effects were observed. The average recovery of analyte incurred in chocolate and baked muffins was < 60 %.

FoodSmartphone, a European project is currently ongoing and proposes the development of smartphone-based (bio)analytical sensing and diagnostics tools for simplified on-site rapid pre-screening of food quality and safety parameters and wireless data transfer to servers of relevant stakeholders (Nielen, 2017). Food allergens are part of the project and the potential to modernize traditional laboratory-based methods by interfacing them with a smartphone readout system was evaluated (Ross *et al*, 2018). Although visionary, this transition from the laboratory to on-site analysis is challenging. Beside analytical aspects, developed methods should be in addition consumer-friendly and adapted to non-scientist users.

DNA-based methods such as PCR are proposed as alternatives to immunological methods. Substantially more stable than proteins, DNA molecules are known to preserve some integrity even under severe food processing conditions. These methods are also very specific and less prone to cross-reactivity phenomena (Costa *et al*, 2016a). However, the choice of the targeted nucleotide sequence is crucial to ensure specificity. Due to high DNA sequence similarities, cross reactivity with soybean was observed by Villa and co-workers in a real-time PCR approach for lupine detection (Villa *et al*, 2018).

Combined with capillary electrophoresis, multiplex PCR was developed to detect multiple food allergen genes in a single reaction (Guo *et al*, 2011). This strategy was implemented by Cheng and co-workers (Cheng *et al*, 2016) in an inter-laboratory study for the simultaneous detection of ten food allergens (hazelnut, pistachio, oat, sesame, peanut, cashew, barley, wheat, soybean and pecan). The method was developed with spiked samples and its applicability was evaluated with 20 commercial food products. This strategy remains questionable since the author conclusions are based on food labelling and precautionary statements. Possible contamination levels of these "real" samples were unknown and obtained results were only qualitative. The use of reference materials with a defined contamination level would have been more appropriate (Mattarozzi & Careri, 2019). Five spiked samples and ten commercial food products were then sent to three other laboratories to perform the same analysis. Results were relatively coherent between laboratories even if cashew was not detected in a sample and hazelnut in two other samples by one of the three laboratories.

However, DNA-based methods suffer from limitations such as the impossibility to differentiate egg from poultry meat and milk from the corresponding meat (Flanagan, 2014). Quantification is also difficult in the absence of certified reference material to convert results expressed in number of copies of DNA equivalent into a total quantity of the allergenic ingredient per kilogram of food.

Mass spectrometry-based methods were recently developed as a promising alternative. Typically, proteins from the allergenic ingredient are detected through their constitutive peptides, which are obtained after proteolytic digestion. The first screening method for the simultaneous detection of multiple allergens was published in 2011 by Heick and co-workers (Heick *et al*, 2011). Since that time, several methods were published for the detection and, based on different strategies, for the quantification of the different allergenic ingredients. A review article of Monaci and co-workers (Monaci *et al*, 2018) identified nine publications presenting multi allergen quantitative mass spectrometry-based methods. For absolute quantification, the common approach requires resorting to stable isotope dilution for the construction matrix matched calibration curves. However, matrix effects rely on sample composition and for the analysis of different kinds of samples, multiple curves would be required. To overcome this issue, inconvenient for routine analysis, Planque and co-workers (Planque *et al*, 2019) developed a quantification strategy based on standard addition.

In these quantitative methods, stable isotope dilution is based on the use of stable isotope-labelled synthetic peptides as internal standard. As already addressed in the introduction, synthetic peptides are not part of all the steps of the sample preparation, as they escape to the enzymatic digestion and are, therefore, not able to correct for the variability introduced at this step of the procedure.

It is in this context that the "Allersens" project was developed. The theses of two PhD students were coordinated to develop and validate a UHPLC-MS/MS method for the quantification of four major allergenic ingredients (egg, milk, peanut and hazelnut) in processed food products. The quantification strategy was based on isotope dilution considering an original approach using a stable isotope-labelled concatemer as internal standard. To achieve this final objective, the project was divided in several

tasks, complementarily addressed by the two students, Kaatje Van Vlierberghe (KVV) and Maxime Gavage (MG).

***1) Production of test materials (KVV)***

To be able to include the impact of food processing in the development and the validation of the analysis method, several test materials, containing the four allergenic ingredients, were produced. These test materials were submitted to several food processing techniques that are known to affect allergen detectability.

***2) Identification of peptide biomarkers (KVV and MG)***

With an empirical approach based on HRMS analysis, potential peptide biomarkers for the development of a quantitative method were identified. The four allergenic ingredients were separately handled and were divided between the two students. KVV was in charge of milk and hazelnut and MG of egg and peanut.

***3) Development of the UHPLC-MS/MS method (KVV)***

The list of HRMS identified potential peptide biomarkers was refined based on selectivity and sensitivity criteria to only keep 19 satisfying peptide biomarkers covering the four allergenic ingredients, in the quantitative method.

***4) Development of a concatemer isotope-labelled internal standard (MG)***

The quantification strategy was based on isotope dilution considering an original approach using a stable isotope-labelled concatemer as internal standard. After several optimization tests, a $^{15}N$ isotope-labelled concatemer, containing the 19 peptide biomarkers, was produced and characterized. This concatemer was used as a unique internal standard in the quantification of the four allergenic ingredients.

***5) Validation of the UHPLC-MS/MS method (KVV and MG)***

The developed UHPLC-MS/MS method was finally validated with a quantification approach based on the use of the $^{15}N$ isotope-labelled concatemer. Performance parameters including selectivity, LOD, LOQ, linearity, trueness and precision were evaluated by both PhD students.

***6) Comparison with existing analysis methods (KVV)***

The performance of the validated UHPLC-MS/MS method will finally be compared with existing allergen analysis methods such as ELISA or PCR but was not covered in this thesis.

Three main tasks were addressed during this thesis: 1) the identification of potential peptide biomarkers for egg and peanut, 2) the development and production of a stable isotope-labelled concatemer internal standard and 3) the validation of the developed UHPLC-MS/MS quantification method.

The identification of the analyte is the first step in the development of any analytical method. In the case of allergen analysis by mass spectrometry, the analytes are constitutive peptides obtained from the enzymatic digestion of proteins of the allergenic ingredients. Potential peptide biomarkers were identified with an empirical approach based on the analysis, by HRMS, of the different allergenic ingredients. Various and representative food processing techniques were applied to these allergenic ingredients to include their impact on allergen detectability. To include potential variability in protein content and protein distribution due to growing conditions and origin, peanuts from two distinct geographical growing locations were considered.

Samples were prepared with an optimized sample preparation protocol and analysed by HPLC-HRMS. With this untargeted approach, hundreds of peptides were identified for egg and peanuts. These peptides were filtered using a set of selection criteria to ensure the specificity, sensitivity and robustness of the future quantitative method. Ideal peptide biomarkers had to be specific to the considered allergenic ingredient, belong to abundant proteins, and be robust to food processing but not prone to missed cleavages by the protease(s) used or to amino acid modifications.

With this approach, 16 potential peptide biomarkers were identified for egg and 16 for peanut. Egg white and yolk both contain allergens and can be separately used by the food industry. Peptide biomarkers from these two fractions were therefore considered. Detailed peptide biomarkers selection was presented in two research articles dedicated to each allergenic ingredient.

The quantification strategy was based on isotope dilution considering an original approach using a stable isotope-labelled concatemer as internal standard. Concatemer is an artificial protein, recombinantly produced and assembling peptide biomarkers from different proteins (Pratt *et al*, 2006). Known for more than a decade in the proteomic field, this strategy has never been applied in allergen analysis or, more broadly, in food analysis. Submitted to enzymatic digestion, concatemer could represent a relevant internal standard in food allergens analysis. Indeed, it overcomes limitations of the use of synthetic peptides, while combining advantages of the use of labelled proteins and, further, allowing for multiple allergen quantification by mass spectrometry.

In parallel to the development of the UHPLC-MS/MS method and the selection of the 19 final peptide biomarkers by KVV, several tests and different concatemers were designed, recombinantly produced, purified and characterized to identify critical design parameters and optimal expression conditions. Aspects such as protein sequence hydrophobicity, associated mRNA secondary structure or *E. coli* host cell strain were investigated.

A concatemer, containing the 19 final peptide biomarkers was finally produced with a $^{15}N$ uniform stable isotope labelling. The obtained production yield made this approach cost-effective compared to synthetic peptides. The isotope labelling strategy and obtained isotopic enrichment were adequate to avoid any risk of false positive introduction.

Performance of the synthetized concatemer was finally compared to those of a labelled protein and of synthetic peptides with the analysis of food matrices spiked with food allergen extracts. As expected from a theoretical point of view, the isotopically labelled protein that was used as an internal standard gave the best results. Synthetic peptides and concatemer results were more equivocal and seemed to be peptide-dependent. These results were gathered in a research article.

The developed UHPLC-MS/MS, targeting 19 peptide biomarkers from the four considered allergenic ingredients, was finally validated. Combining the use of a $^{15}N$ isotope-labelled concatemer internal standard, incurred and spiked food matrices and standard addition quantification, performance parameters including selectivity, LOD, LOQ, linearity, trueness and precision were evaluated. The

results were compared to the standard method performance required by the AOAC International for the detection and the quantification of selected food allergens. All requirements were not achieved. Sensitivity issues were observed for egg and in complex matrices such as chocolate, known to affect sample preparation. However, linearity was observed and at least one peptide per allergenic ingredient was in agreement with AOAC requirements concerning trueness and precision.

The proposed strategy and the developed quantitative method are major steps forward in allergen analysis. Identified related literature rarely considers all aspects involved in food allergen analysis. Here, a quantitative method using an original isotope dilution strategy was developed. This method targeted multiple allergenic ingredients, include food processing effects and was validated. These three aspects are crucial for the valuable expansion of mass spectrometry in food allergen analysis.

Presented results were limited to a single laboratory validation. However, as a first perspective, an inter-laboratory study is planned in a near future, during the last months of the "Allersens" project (KVV). More than ten laboratories were identified, based on voluntary approach, to participate to this inter-laboratory study. The sample preparation protocol, contaminated samples, $^{15}N$ isotope-labelled concatemer internal standard and UHPLC-MS/MS method will be provided to all participants. The developed quantitative method will therefore be applied in several laboratories and information concerning method reproducibility will be collected. In a more utopian perspective, this inter-laboratory study could be the first step towards method harmonisation between the different control laboratories. This harmonisation is crucial to obtain comparable results among laboratories and to promote an efficient food allergens risk management.

Compared to AOAC International standard method requirements, obtained results were promising. A substantial part of the performance criteria were fulfilled. The main identified weak point concerned method sensitivity, for the detection of egg in general on the one hand and detection in complex matrices such as chocolate on the other hand. Efforts have to be focused on sample preparation to improve sensitivity. However, this task is challenging in the context of routine laboratories. Any proposed improvement has to be in accordance with routine laboratory imperatives in terms of analysis cost and duration.

It was, for example, demonstrated in the research article on the identification of potential egg peptide biomarkers by HRMS (Gavage *et al*, 2019) that the use of Lys-C in combination with trypsin for the enzymatic digestion step could improve the detection of some proteins such as ovalbumin. However, this additional step elongates the duration of the digestion step and is an additional cost. For practical reasons, the strategy of the laboratory is to use the same sample preparation protocol, independently of the researched allergenic ingredient or the type of matrix. Therefore, one can argue about whether the addition of Lys-C in the digestion step of every sample is justified to improve the detection of only one allergenic ingredient.

For chocolate matrices, one could suggest to add a purification step after the extraction to separate protein from other matrix components, such as polyphenols and tannins, which are known to negatively affect the following enzymatic digestion step. This separation could be achieved using size-exclusion chromatographic disposable columns or immunoaffinity chromatography. But, once again, the addition of such a step is questionable. Does this step have to be applied to every sample? If it does, it could improve allergen detection sensitivity in complex matrices but a part of the extracted protein would inevitably be lost during this step, which would be detrimental to simpler matrices. If not, on which basis does the choice of the additional purification step have to be done? What if a cookie containing chocolate chips would have to be analysed? The purification step would be beneficial for the chocolate matrix but not necessarily for the cookie matrix.

Method sensitivity could definitely be improved but the cost/benefit ratio has to be carefully evaluated to ensure method financial sustainability in the context of routine laboratories.

As exposed in the research article published in *Food Chemistry*, the performance of the designed concatemer was evaluated and compared to isotope-labelled protein and synthetic peptides. The analysis of three food matrices spiked with extracts of four allergenic ingredients indicated that, as expected from a theoretical point of view, the isotope-labelled protein was the most performing internal standard, even if results were limited to β-lactoglobulin, a milk protein. Results obtained for synthetic peptides and the concatemer were somewhat comparable and were depending on the targeted peptide biomarker.

The design of the $^{15}N$ isotope-labelled concatemer was optimized regarding production factors. The 19 peptide biomarkers were simply concatenated in the concatemer sequence. However, multiple factors could influence the enzymatic digestion kinetics and different possibilities of improvement were proposed.

For instance, amino acids surrounding trypsin recognition sites are known to influence the efficiency of peptide bond hydrolysis (Siepen *et al*, 2007). This aspect was already considered during the potential peptide biomarkers selection by HRMS. Peptides with an arginine or lysine residue juxtaposed to the digestion site were rejected. However, due to peptides concatenation, each cleavage site close environment corresponds to the half of the cleavage site of the natural protein. Consequently, at a local scale, enzymatic digestion of the concatemer only partially reflects digestion of the natural proteins. The introduction of amino acids between each targeted peptide of the concatemer could be a solution to overcome this issue (Kito *et al*, 2007). Such introduced amino acids would be the flanking amino acids in the corresponding natural protein sequence.

Protein structural parameters are also known to interfere with trypsin digestion. Amino acids surrounding the cleavage site in the three-dimensional structure of the protein might also affect trypsin digestion. Hence, cleavage sites surrounded by acidic amino acids, characterized by a greater average exposed area, are more subject to missed-cleavages. Moreover, protein secondary structure was found to affect trypsin digestion efficiency. Cleavage sites within unstructured domains are more prone to be cleaved incorrectly, whereas cleavage sites in alpha-helices are more favourable (Hamady *et al*, 2005). Even if no general trend emerged from our data, three-dimensional and structural aspects could be included in a future peptide biomarker selection, in addition to all other criteria.

In this project, four allergenic ingredients were considered. Egg, milk, peanut and hazelnut were identified as first priority food allergens based on their prevalence, their allergic potency and the severity of the allergic reaction (incidence of effects). The developed strategy could now be extended to other allergenic ingredients that require labelling and that are listed in Annex II of Consumer Information Regulation (EC) 1169/2011 (Table 1). Except for sulphur dioxide, all these ingredients contain proteins whose peptides can be targets of the method. After the identification by HRMS of potential peptide biomarkers that fulfil selection criteria (specificity for the allergenic ingredient, derived from an abundant protein, robust to food processing and not subjected to amino acid modifications or missed enzymatic cleavages), these peptides would be included in another concatemer used as stable isotope-labelled internal standard for the quantification of multiple food allergens. Considering multiple peptides by allergenic ingredient, a single concatemer for the analysis of the 13 allergenic ingredients that required labelling (sulphur dioxide excluded) could be laborious and not strategic. It could be considered to produce a couple of stable isotope-labelled concatemers and to group in each concatemer, peptides from allergenic ingredients for which the analysis request

is often clustered such as peanut and tree nuts (almonds, hazelnuts, walnuts, cashews, pecan nuts, Brazil nuts, pistachio nuts and macadamia) or fish and crustacean, for example.

Moreover, the list of 14 allergenic ingredients and substances causing intolerances that require labelling is probably not definitive. Lupine and molluscs were only added to this list in the last version of the Regulation in 2011 (Popping & Diaz-Amigo, 2018). This list will possibly evolve in the future and other allergenic ingredients could be added, such as fruits like kiwifruit or banana which are already included in the Japanese food allergen labelling legislation (Akiyama *et al*, 2011). Allergy to kiwifruit was first described in 1981. Since then, there has been an increasing number of reports of kiwifruit allergy. Today it is one of the most common causes of food allergy (Le *et al*, 2013). Symptoms of kiwifruit allergy vary from mild local symptoms in the oral cavity to severe systemic reactions (Lucas *et al*, 2003). Avoidance is challenging since kiwifruit is a versatile ingredient, which is found in various food products worldwide including cakes, juice, jam, ice cream, fruit wine, or fruit salads (Wang *et al*, 2019). This list could also be subject to modifications due to the introduction of novel foods for human consumption. Insects represent an alternative for meat and fish in satisfying the increasing demand for sustainable sources of nutrition. Food allergy to insects has already been described for several edible species such as mealworm or cricket (de Gier & Verhoeckx, 2018). Analytical methods will therefore be necessary for the detection of these potentially new allergenic ingredients that required labelling. A couple of detection methods are already available in the literature such as the work of Zhang and co-workers for the development of sandwich ELISA for detection of invertebrate major allergen tropomyosin (Zhang *et al*, 2014a) or the work of Suh and co-workers for the development of a multiplex PCR method for the simultaneous detection of fruit allergen-coding genes in tomato, apple, peach and kiwifruit (Suh *et al*, 2019).

The analytical strategy developed in this project can even be extended to other fields using proteomic analysis. The use of biological systems to synthesize recombinant proteins as therapeutic products has been a remarkable success since 1982 and the approval by the U.S. Food and Drug Administration of Humulin, the first human insulin for diabetes treatment created using recombinant DNA technology (De Meyts, 2017). In recent years, antibody-drug conjugates, a new class of therapeutics, have been developed. In these cases, the protein counterpart acts as a targeted vehicle for conventional chemical drugs (Sanchez-Garcia *et al*, 2016). In each case, the product must be purified from any cell-based impurities to an acceptable level before administration in the clinic. Host cell proteins are one of the more troublesome impurities in terms of risk or product degradation, for example, residual proteases that could potentially reduce the amount of active drug present (Bracewell *et al*, 2015). Such kind of therapeutic products therefore need control analysis targeting these host cell proteins. The strategy developed in this thesis for the detection of multiple allergens in processed food products, based on the targeted analysis of peptide biomarkers by UHPLC-MS/MS and the use of an isotope labelled concatemer as internal standard, could be transferred for the analysis of multiple host cell proteins in therapeutic products.

A UHPLC-MS/MS based method targeting signature tryptic peptides was already used to detect blood-derived products and milk powder in animal feed (Lecrenier *et al*, 2018). This method resulted from a collaboration between the universities of Namur and Liège, the Walloon Agricultural Research Centre and CER Groupe. This method was developed in the context of the use of animal by-products as a source of proteins in the production of animal feed. Since the bovine spongiform encephalopathy crisis ("mad cow disease"), the use of animal by-products in feed is strictly regulated. In the studied case of fish feed, the use of non-ruminant processed animal proteins (PAPs) is authorized as well as ruminant milk. On the contrary, the use of ruminant PAPs is forbidden.

A UHPLC-MS/MS method was developed to overcome the limitations of DNA-based methods which are unable to differentiate milk from ruminant PAPs. The method targeted signature peptides from bovine blood and milk proteins to simultaneously identify authorised and unauthorised animal by-products. A stable isotope-labelled internal standard was used in this method under the form of two synthetic peptides, each containing two concatenated peptide biomarkers. This approach was somehow between synthetic peptides and recombinantly produced concatemer.

A new $^{15}N$ stable isotope-labelled concatemer internal standard is currently under development to improve this feed UHPLC-MS/MS analysis method. This concatemer contains 14 peptide biomarkers from different milk and animal by-products proteins. Its design includes the introduction of flanking amino acids in the corresponding natural protein sequence between each peptide biomarker. With this additional feature, the aim is to improve the similarity between analyte and internal standard behaviour during the enzymatic digestion. We are looking forwards to the results.

In conclusion, these examples demonstrate the potential opportunities of extension of the developed strategy. Applied to food allergens, a UHPLC-MS/MS method was developed and validated for the simultaneous quantification of egg, milk, peanut and hazelnut in processed food products. Food allergen analysis is an essential tool in the development of a risk-based approach to allergen management to provide safe food and a better living conditions to allergic patients.

**REFERENCES**


Abernathy-Carver KJ, Sampson HA, Picker LJ & Leung DYM (1995) Milk-induced eczema is associated with the expansion of T cells expressing cutaneous lymphocyte antigen. *J. Clin. Invest.* **95:** 913–918

Addgene Addgene: Vector Database - pET-17b. Available at: https://www.addgene.org/vector-database/2545/ [Accessed April 3, 2020]

Akiyama H, Imai T & Ebisawa M (2011) Japan Food Allergen Labeling Regulation-History and Evaluation. In *Advances in Food and Nutrition Research* pp 139–171.

Albillos SM, Al-Taher F & Maks N (2011) Increasing extractability of protein for allergen detection after food processing. *Food Chem.* **127:** 1831–1834

Allen KJ, Turner PJ, Pawankar R, Taylor S, Sicherer S, Lack G, Rosario N, Ebisawa M, Wong G, Mills ENC, Beyer K, Fiocchi A & Sampson HA (2014) Precautionary labelling of foods for allergen content: Are we ready for a global framework? *World Allergy Organ. J.* **7:** 1–14

Alvarez PA & Boye JI (2012) Food Production and Processing Considerations of Allergenic Food Ingredients: A Review. *J. Allergy* **2012:** 14

Andjelković U, Gavrović-Jankulović M, Martinović T & Josić D (2017) Omics methods as a tool for investigation of food allergies. *TrAC - Trends Anal. Chem.* **96:** 107–115

Apostolovic D, Stanic-Vucinic D, de Jongh HHJ, de Jong GAH, Mihailovic J, Radosavljevic J, Radibratovic M, Nordlee JA, Baumert JL, Milcic M, Taylor SL, Garrido Clua N, Cirkovic Velickovic T & Koppelman SJ (2016) Conformational stability of digestion-resistant peptides of peanut conglutins reveals the molecular basis of their allergenicity. *Sci. Rep.* **6:** 1–12

Arshad SH (2016) An update on allergen immunotherapy. *Clin. Med. J. R. Coll. Physicians London* **16:** 584

Baumert JL (2014) Detecting and Measuring Allergens in Food. In *Risk Management for Food Allergy* pp 215–226. Academic Press

Bender ML, Killheffer J V. & Cohen S (1973) Chymotrypsin. *Crit. Rev. Biochem. Mol. Biol.* **1:** 149–199

Berthold DA, Jeisy VJ, Sasser TL, Shea JJ, Frericks HL, Shah G & Rienstra CM (2011) Top ten tips for producing 13C, 15N protein in abundance

Beyer K (2012) A European perspective on immunotherapy for food allergies. *J. Allergy Clin. Immunol.* **129:** 1179–1184

BioLabs NE (2017) Instruction Manual Q5 ® Site-Directed Mutagenesis Kit

Black KR, Filep S, Smith BRE, Lee J, Murphy B, Block D & Chapman MD (2019) Simultaneous Quantification of Major Food Allergens Using Fluorescent Multiplex Array. *J. Allergy Clin. Immunol.* **143:** AB240

Bornhorst JA & Falke JJ (2000) Purification of proteins using polyhistidine affinity tags. *Methods Enzymol.* **326:** 245–254

Bracewell DG, Francis R & Smales CM (2015) The future of host cell protein (HCP) identification during process development and manufacturing linked to a risk-based management for their control. *Biotechnol. Bioeng.* **112:** 1727–1737

Brežná B, Hudecová L & Kuchta T (2006) A novel real-time polymerase chain reaction (PCR) method for the detection of walnuts in food. *Eur. Food Res. Technol.* **223:** 373–377

Brough HA, Simpson A, Makinson K, Hankinson J, Brown S, Douiri A, Belgrave DCM, Penagos M, Stephens AC, McLean WHI, Turcanu V, Nicolaou N, Custovic A & Lack G (2014) Peanut allergy: Effect of environmental peanut exposure in children with filaggrin loss-of-function mutations. *J. Allergy Clin. Immunol.* **134:** 867–875

Brownridge PJ, Harman VM, Simpson DM & Beynon RJ (2012) Absolute multiplexed protein quantification using qconcat technology. In *Quantitative Methods in Proteomics. Methods in Molecular Biology (Methods and Protocols)* pp 267–293.

Brunner K (2016) Beyond Immuno-Based Allergen Testing. Available at: https://www.romerlabs.com/fr/centre-d-expertise/base-documentaire/articles/news/beyond-immuno-based-allergen-testing/ [Accessed February 24, 2020]

Budhathoki S (2018) ELISA-Types,Applications | Online notes on microbiology. Available at: https://microbionotes.com/elisa-typesapplications/ [Accessed February 24, 2020]

Bunkenborg J, Espadas G & Molina H (2013) Cutting edge proteomics: Benchmarking of six commercial trypsins. *J. Proteome Res.* **12:** 3631–3641

Burkhart JM, Schumbrutzki C, Wortelkamp S, Sickmann A & Zahedi RP (2012) Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics. *J. Proteomics* **75:** 1454–1462

Del Campo C, Bartholomäus A, Fedyunin I & Ignatova Z (2015) Secondary Structure across the Bacterial Transcriptome Reveals Versatile Roles in mRNA Regulation and Function. *PLoS Genet.* **11:** e1005613

Carrard A, Rizzuti D & Sokollik C (2015) Update on food allergy. *Allergy Eur. J. Allergy Clin. Immunol.* **70:** 1511–1520

Chawner R, McCullough B, Giles K, Barran PE, Gaskell SJ & Eyers CE (2012) QconCAT standard for calibration of ion mobility-mass spectrometry systems. *J. Proteome Res.* **11:** 5564–5572

Chen H, Bjerknes M, Kumar R & Jay E (1994) Determination of the optimal aligned spacing between the shine - dalgarno sequence and the translation initiation codon of escherichia coli m RNAs. *Nucleic Acids Res.* **22:** 4953–4957

Chen Q, Zhang J, Ke X, Lai S, Li D, Yang J, Mo W & Ren Y (2016) Simultaneous quantification of α-lactalbumin and β-casein in human milk using ultra-performance liquid chromatography with tandem mass spectrometry based on their signature peptides and winged isotope internal standards. *Biochim. Biophys. Acta - Proteins Proteomics* **1864:** 1122–1127

Cheng F, Wu J, Zhang J, Pan A, Quan S, Zhang D, Kim HY, Li X, Zhou S & Yang L (2016) Development and inter-laboratory transfer of a decaplex polymerase chain reaction assay combined with capillary electrophoresis for the simultaneous detection of ten food allergens. *Food Chem.* **199:** 799–808

Cho CY, Nowatzke W, Oliver K & Garber EAE (2015) Multiplex detection of food allergens and gluten. *Anal. Bioanal. Chem.* **407:** 4195–4206

Cianferoni A & Spergel JM (2009) Food allergy: Review, classification and diagnosis. *Allergol. Int.* **58:** 457–466

Clemente A, Chambers SJ, Lodi F, Nicoletti C & Brett GM (2004) Use of the indirect competitive ELISA for the detection of Brazil nut in food products. *Food Control* **15:** 65–69

Commins SP (2015) Mechanisms of Oral Tolerance. *Pediatr. Clin. North Am.* **62:** 1523–1529

Corthay A (2009) How do regulatory t cells work? *Scand. J. Immunol.* **70:** 326–336

Costa J, Fernandes TJR, Villa C, Oliveira MBPP & Mafra I (2016a) Advances in food allergen analysis. In *Food Safety: Innovative Analytical Tools for Safety Assessment* pp 305–360.

Costa J, Mafra I, Carrapatoso I & Oliveira MBPP (2016b) Hazelnut allergens: Molecular characterization, detection, and clinical relevance. *Crit. Rev. Food Sci. Nutr.* **56:** 2579–2605

Croote D & Quake SR (2016) Food allergen detection by mass spectrometry: The role of systems biology. *npj Syst. Biol. Appl.* **2:** 1–10

DunnGalvin A, Chan CH, Crevel R, Grimshaw K, Poms R, Schnadt S, Taylor SL, Turner P, Allen KJ, Austin M, Baka A, Baumert JL, Baumgartner S, Beyer K, Bucchini L, Fernández-Rivas M, Grinter K, Houben GF, Hourihane J, Kenna F, et al (2015) Precautionary allergen labelling: Perspectives from key stakeholder groups. *Allergy Eur. J. Allergy Clin. Immunol.* **70:** 1039–1051

Eckl-Dorna J, Villazala-Merino S, Linhart B, Karaulov A V., Zhernov Y, Khaitov M, Niederberger-Leppin V & Valenta R (2019) Allergen-specific antibodies regulate secondary allergen-specific immune responses. *Front. Immunol.* **9:**

EFSA (2014) Scientific Opinion on the evaluation of allergenic foods and food ingredients for labelling purposes. *EFSA J.* **12:** 3894

Eggesbø M, Botten G, Halvorsen R & Magnus P (2001) The prevalence of allergy to egg: A population-based study in young children. *Allergy Eur. J. Allergy Clin. Immunol.* **56:** 403–411

Ehlert A, Demmel A, Hupfer C, Busch U & Engel KH (2009) Simultaneous detection of DNA from 10 food allergens by ligation-dependent probe amplification. *Food Addit. Contam. - Part A Chem. Anal. Control. Expo. Risk Assess.* **26:** 409–418

Fearn T & Thompson M (2001) A new test for 'sufficient homogeneity'. *Analyst* **126:** 1414–1417

Fellows PJ (2009) Food Processing Technology: Principles and Practice Elsevier Science Available at: https://books.google.be/books?id=t4ykAgAAQBAJ

Filipp F V., Sinha N, Jairam L, Bradley J & Opella SJ (2009) Labeling strategies for 13C-detected aligned-sample solid-state NMR of proteins. *J. Magn. Reson.* **201:** 121–130

Flanagan S (2014) Handbook of Food Allergen Detection and Control. Elsevier Science

Frank S (2002) Specificity and Cross-Reactivity. In *Immunology and Evolution of Infectious Disease* pp 33–56.

Fritsch R, Bohle B, Vollmann U, Wiedermann U, Jahn-Schmid B, Krebitz M, Breiteneder H, Kraft D & Ebner C (1998) Bet v 1, the major birch pollen allergen, and Mal d 1, the major apple allergen, cross-react at the level of allergen- Specific T helper cells. *J. Allergy Clin. Immunol.* **102:** 679–686

Frottin F, Martinez A, Peynot P, Mitra S, Holz RC, Giglione C & Meinnel T (2006) The proteomics of N-terminal methionine cleavage. *Mol. Cell. Proteomics* **5:** 2336–2349

Fukutomi Y, Taniguchi M, Nakamura H & Akiyama K (2014) Epidemiological link between wheat allergy and exposure to hydrolyzed wheat protein in facial soap. *Allergy Eur. J. Allergy Clin. Immunol.* **69:** 1405–1411

Gaspar P, Moura G, Santos MAS & Oliveira JL (2013) MRNA secondary structure optimization using a correlated stem-loop prediction. *Nucleic Acids Res.* **44:** 5490

Gavage M, Van Vlierberghe K, Van Poucke C, De Loose M, Gevaert K, Dieu M, Renard P, Arnould T, Filee P & Gillard N (2020a) Comparative study of concatemer efficiency as an isotope-labelled

internal standard for allergen quantification. *Food Chem.* **332:** 127413

Gavage M, Van Vlierberghe K, Van Poucke C, De Loose M, Gevaert K, Dieu M, Renard P, Arnould T & Gillard N (2019) Selection of egg peptide biomarkers in processed food products by high resolution mass spectrometry. *J. Chromatogr. A* **1584:** 115–125

Gavage M, Van Vlierberghe K, Van Poucke C, De Loose M, Gevaert K, Dieu M, Renard P, Arnould T & Gillard N (2020b) High-resolution mass spectrometry-based selection of peanut peptide biomarkers considering food processing and market type variation. *Food Chem.* **304:** 125428

Gendel S, Buchanan R, Dennis S, Acheson D, Assimon SA, Beru N, Bolger P, Carlson D, Carvajal R, Copp C, Falci K, Garber E, Harden E, Kane R, Kvenberg J, Luccioli S, Park D, Raybourne R, Troxell T & Vierk K (2008) Approaches to establish thresholds for major food allergens and for gluten in food. *J. Food Prot.* **71:** 1043–1088

de Gier S & Verhoeckx K (2018) Insect (food) allergy and allergens. *Mol. Immunol.* **100:** 82–106

Glatter T, Ludwig C, Ahrné E, Aebersold R, Heck AJR & Schmidt A (2012) Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. *J. Proteome Res.* **11:** 5145–5156

Gomaa A & Boye JI (2013) Impact of thermal processing time and cookie size on the detection of casein, egg, gluten and soy allergens in food. *Food Res. Int.* **52:** 483–489

González AG, Herrador MÁ & Asuero AG (2010) Intra-laboratory assessment of method accuracy (trueness and precision) by using validation standards. *Talanta* **82:** 1995–1998

Gorochowski TE, Ignatova Z, Bovenberg RAL & Roubos JA (2015) Trade-offs between tRNA abundance and mRNA secondary structure support smoothing of translation elongation rate. *Nucleic Acids Res.* **43:** 3022–3032

Green PHR, Lebwohl B & Greywoode R (2015) Celiac disease. *J. Allergy Clin. Immunol.* **135:** 1099–1106

Guo J, Yang L, Chen L, Morisset D, Li X, Pan L & Zhang D (2011) MPIC: A high-throughput analytical method for multiple DNA targets. *Anal. Chem.* **83:** 1579–1586

Gupta RS, Springston EE, Warrier MR, Smith B, Kumar R, Pongracic J & Holl JL (2011) The prevalence, severity, and distribution of childhood food allergy in the United States. *Pediatrics* **128:** e9–e17

Gupta RS, Taylor SL, Baumert JL, Kao LM, Schuster E & Smith BM (2017) Economic factors impacting food allergen management: Perspectives from the food industry. *J. Food Prot.* **80:** 1719–1725

Hamady M, Cheung THT, Tufo H & Knight R (2005) Does protein structure influence trypsin miscleavage? *IEEE Eng. Med. Biol. Mag.* **24:** 58–66

Hawkes D (2019) The allergen bureau-helping the food industry navigate allergen science and manage risk. *Food New Zeal.* **19:** 19

Hayder H, Mueller U & Bartholomaeus A (2011) Review of Intolerance Reactions to Food and Food Additives. *Int. Food Risk Anal. J.* **1:** 23–32

He SH, Zhang HY, Zeng XN, Chen D & Yang PC (2013) Mast cells and basophils are essential for allergies: Mechanisms of allergic inflammation and a proposed procedure for diagnosis. *Acta Pharmacol. Sin.* **34:** 1270–1283

Heick J, Fischer M & Pöpping B (2011) First screening method for the simultaneous detection of seven allergens by liquid chromatography mass spectrometry. *J. Chromatogr. A* **1218:** 938–943

Herman RA, Scherer PN & Shan G (2008) Evaluation of logistic and polynomial models for fitting sandwich-ELISA calibration curves. *J. Immunol. Methods* **339:** 245–258

Ho MHK, Wong WHS & Chang C (2014) Clinical spectrum of food allergies: A comprehensive review. *Clin. Rev. Allergy Immunol.* **46:** 225–240

Holtug K, Clausen MR, Hove H, Christiansen J & Mortensen PB (1992) The colon in carbohydrate malabsorption: Short-chain fatty acids, ph, and osmotic diarrhoea. *Scand. J. Gastroenterol.* **27:** 545–552

Holzhauser T & Röder M (2015) Polymerase chain reaction (PCR) methods for detecting allergens in foods. In *Handbook of Food Allergen Detection and Control* pp 245–263.

Holzhauser T, Stephan O & Vieths S (2002) Detection of potentially allergenic hazelnut (Corylus avellana) residues in food: A comparative study with DNA PCR-ELISA and protein sandwich-ELISA. *J. Agric. Food Chem.* **50:** 5808–5815

Horwitz W, Kamps LR & Boyer KW (1980) Quality assurance in the analysis of foods and trace constituents. *J. Assoc. Off. Anal. Chem.* **63:** 1344–1354

Houser B (2012) Bio-rad's Bio-Plex® suspension array system, xMAP technology overview. *Arch. Physiol. Biochem.* **118:** 192–196

Hupfer C, Waiblinger HU & Busch U (2007) Development and validation of a real-time PCR detection method for celery in food. *Eur. Food Res. Technol.* **225:** 329–335

Husain Z & Schwartz RA (2012) Peanut allergy: An increasingly common life-threatening disorder. *J. Am. Acad. Dermatol.* **66:** 136–143

Husby S, Foged N, Høst A & Svehag SE (1987) Alimentary tract and pancreas passage of dietary antigens into the blood of children with coeliac disease. Quantification and size distribution of absorbed antigens. *Gut* **28:** 1062–1072

Huynh ML, Russell P & Walsh B (2009) Tryptic digestion of in-gel proteins for mass spectrometry analysis. In *Methods in molecular biology (Clifton, N.J.)* pp 507–513.

Invitrogen (2010) One Shot BL21 Star (DE3) and One Shot BL21 Star (DE3) pLysS Competent Cells. Available at: https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2Foneshotbl21star_man.pdf&title=T25lIFNob3QgQkwyMSBTdGFyIChERTMpIGFuZCBPbmUgU2hvdCBCTDIxIFN0YXIgKERFMykgcEx5c1MgQ29t [Accessed April 3, 2020]

Iqbal A, Shah F, Hamayun M, Ahmad A, Hussain A, Waqas M, Kang SM & Lee IJ (2016) Allergens of Arachis hypogaeaand the effect of processing on their detection by ELISA. *Food Nutr. Res.* **60:** 28945

Issa K (2013) Effet protecteur du sulfure d'hydrogène, de la protéine C activée et de la dexamétasone dans la modulation hémodynamique et inflammatoire de l'ischémie/reperfusion.

James JM (2003) Respiratory manifestations of food allergy. *Pediatrics* **111:** 1625–1630

Jana A (2017) High Protein Milk Ingredients - A Tool for Value-Addition to Dairy and Food Products. *J. Dairy, Vet. Anim. Res.* **6:** 171

Jo J, Garssen J, Knippels L & Sandalova E (2014) Role of cellular immunity in cow's milk allergy: Pathogenesis, tolerance induction, and beyond. *Mediators Inflamm.* **2014:** 10

Johansson SGO, Bieber T, Dahl R, Friedmann PS, Lanier BQ, Lockey RF, Motala C, Ortega Martell JA,

Platts-Mills TAE, Ring J, Thien F, Van Cauwenberge P & Williams HC (2004) Revised nomenclature for allergy for global use: Report of the Nomenclature Review Committee of the World Allergy Organization, October 2003. *J. Allergy Clin. Immunol.* **113:** 832–836

Johnson PE, Baumgartner S, Aldick T, Bessant C, Giosafatto V, Heick J, Mamone G, O'Connor G, Poms R, Popping B, Reuter A, Ulberth F, Watson A, Monaci L & Mills ENC (2011) Current perspectives and recommendations for the development of mass spectrometry methods for the determination of allergens in foods. *J. AOAC Int.* **94:** 1026–1033

Joint FAO/WHO Food Standards Programme (2007) Codex Alimentarius Commission Procedural Manual. In *Codex Alimentarius Commission*

Kim JS, Monroe ME, Camp DG, Smith RD & Qian WJ (2013) In-source fragmentation and the sources of partially tryptic peptides in shotgun proteomics. *J. Proteome Res.* **12:** 910–916

Kirsch S, Fourdrilis S, Dobson R, Scippo ML, Maghuin-Rogister G & De Pauw E (2009) Quantitative methods for food allergens: A review. *Anal. Bioanal. Chem.* **395:** 57–67

Kito K, Ota K, Fujita T & Ito T (2007) A synthetic protein approach toward accurate mass spectrometric quantification of component stoichiometry of multiprotein complexes. *J. Proteome Res.* **6:** 792–800

Koeberl M, Clarke D & Lopata AL (2014) Next generation of food allergen quantification using mass spectrometric systems. *J. Proteome Res.* **13:** 3499–3509

Koeberl M, Sharp MF, Tian R, Buddhadasa S, Clarke D & Roberts J (2018) Lupine allergen detecting capability and cross-reactivity of related legumes by ELISA. *Food Chem.* **256:** 105–112

Koplin JJ, Allen KJ, Gurrin LC, Peters RL, Lowe AJ, Tang MLK, Dharmage SC, Ponsonby AL, Hill D, Matheson M, Wake M, Thiele L, Czech H, Eckert J, Anderson D, Hamilton O, Bertalli N, Sanjeevan J, Dang T, Tan T, et al (2013) The impact of family history of allergy on risk of food allergy: A population-based study of infants. *Int. J. Environ. Res. Public Health* **10:** 5364–5377

Koplin JJ, Osborne NJ & Allen KJ (2010) Prevalence of allergen avoidance advisory statements on packaged processed foods in a supermarket. *Med. J. Aust.* **193:** 426

Köppel R, Dvorak V, Zimmerli F, Breitenmoser A, Eugster A & Waiblinger HU (2010) Two tetraplex real-time PCR for the detection and quantification of DNA from eight allergens in food. *Eur. Food Res. Technol.* **230:** 367

Koppelman SJ, Hefle SL, Taylor SL & De Jong GAH (2010) Digestion of peanut allergens Ara h 1, Ara h 2, Ara h 3, and Ara h 6: A comparative in vitro study and partial characterization of digestion-resistant peptides. *Mol. Nutr. Food Res.* **54:** 1711–1721

Koppelman SJ, Knol EF, Vlooswijk RAA, Wensing M, Knulst AC, Hefle SL, Gruppen H & Piersma S (2003) Original article Peanut allergen Ara h 3 : Isolation from peanuts and biochemical characterization. *Allergy* **58:** 1144–1151

Korte R, Oberleitner D & Brockmeyer J (2019) Determination of food allergens by LC-MS: Impacts of sample preparation, food matrix, and thermal processing on peptide detectability and quantification. *J. Proteomics* **196:** 131–140

Kushnir MM, Rockwood AL, Nelson GJ, Yue B & Urry FM (2005) Assessing analytical specificity in quantitative analysis using tandem mass spectrometry. *Clin. Biochem.* **38:** 319–327

Kyte J & Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157:** 105–132

Lack G, Fox D, Northstone K & Golding J (2003) Factors associated with the development of peanut allergy in childhood. *N. Engl. J. Med.* **348:** 977–985

Le TM, Bublin M, Breiteneder H, Fernández-Rivas M, Asero R, Ballmer-Weber B, Barreales L, Bures P, Belohlavkova S, De Blay F, Clausen M, Dubakiene R, Gislason D, Van Hoffen E, Jedrzejczak-Czechowicz M, Kowalski ML, Kralimarkova T, Lidholm J, Dewitt ÅM, Mills CEN, et al (2013) Kiwifruit allergy across Europe: Clinical manifestation and IgE recognition patterns to kiwifruit allergens. *J. Allergy Clin. Immunol.* **131:** 164–171

Lecrenier MC, Planque M, Dieu M, Veys P, Saegerman C, Gillard N & Baeten V (2018) A mass spectrometry method for sensitive, specific and simultaneous detection of bovine blood meal, blood products and milk products in compound feed. *Food Chem.* **245:** 981–988

Lin Y, Zhou J, Bi D, Chen P, Wang X & Liang S (2008) Sodium-deoxycholate-assisted tryptic digestion and identification of proteolytically resistant proteins. *Anal. Biochem.* **377:** 259–266

Linacero R, Ballesteros I, Sanchiz A, Prieto N, Iniesto E, Martinez Y, Pedrosa MM, Muzquiz M, Cabanillas B, Rovira M, Burbano C & Cuadrado C (2016) Detection by real time PCR of walnut allergen coding sequences in processed foods. *Food Chem.* **202:** 334–340

Linacero R, Sanchiz A, Ballesteros I & Cuadrado C (2020) Application of real-time PCR for tree nut allergen detection in processed foods. *Crit. Rev. Food Sci. Nutr.* **60:** 1077–1093

Liu AH, Jaramillo R, Sicherer SH, Wood RA, Bock SA, Burks AW, Massing M, Cohn RD & Zeldin DC (2010) National prevalence and risk factors for food allergy and relationship to asthma: Results from the National Health and Nutrition Examination Survey 2005-2006. *J. Allergy Clin. Immunol.* **126:** 798–806

Loh W & Tang MLK (2018) The epidemiology of food allergy in the global context. *Int. J. Environ. Res. Public Health* **15:** 2043

Lucas JSA, Lewis SA & Hourihane JOB (2003) Kiwi fruit allergy: A review. *Pediatr. Allergy Immunol.* **14:** 420–428

Lund MN & Ray CA (2017) Control of Maillard Reactions in Foods: Strategies and Chemical Mechanisms. *J. Agric. Food Chem.* **65:** 4537–4552

MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC & MacCoss MJ (2010) Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26:** 966–968

Mali S & Jambure R (2012) Anaphyllaxis management: Current concepts. *Anesth. Essays Res.* **6:** 115

Marenholz I, Grosche S, Kalb B, Rüschendorf F, Blümchen K, Schlags R, Harandi N, Price M, Hansen G, Seidenberg J, Röblitz H, Yürek S, Tschirner S, Hong X, Wang X, Homuth G, Schmidt CO, Nöthen MM, Hübner N, Niggemann B, et al (2017) Genome-wide association study identifies the SERPINB gene cluster as a susceptibility locus for food allergy. *Nat. Commun.* **8:** 1056 Available at: https://doi.org/10.1038/s41467-017-01220-0

Martino DJ & Prescott SL (2010) Silent mysteries: Epigenetic paradigms could hold the key to conquering the epidemic of allergy and immune disease. *Allergy Eur. J. Allergy Clin. Immunol.* **65:** 7–15

Mattarozzi M, Bignardi C, Elviri L & Careri M (2012) Rapid shotgun proteomic liquid chromatography-electrospray ionization-tandem mass spectrometry-based method for the lupin (Lupinus albus L.) multi-allergen determination in foods. *J. Agric. Food Chem.* **60:** 5841–5846

Mattarozzi M & Careri M (2019) The role of incurred materials in method development and

validation to account for food processing effects in food allergen analysis. *Anal. Bioanal. Chem.* **441:** 4465–4480

McLean-Tooke APC, Bethune CA, Fay AC & Spickett GP (2003) Adrenaline in the treatment of anaphylaxis: What is the evidence? *Br. Med. J.* **327:** 1332–1335

McWilliam V, Koplin J, Lodge C, Tang M, Dharmage S & Allen K (2015) The Prevalence of Tree Nut Allergy: A Systematic Review. *Curr. Allergy Asthma Rep.* **15:** 54

De Meyts P (2017) Early Recombinant Protein Therapeutics. *Protein Ther.***:** 1–23 Available at: https://doi.org/10.1002/9783527699124.ch1

Monaci L, De Angelis E, Montemurro N & Pilolli R (2018) Comprehensive overview and recent advances in proteomics MS based methods for food allergens analysis. *TrAC - Trends Anal. Chem.* **106:** 21–36

Monaci L & Van Hengel AJ (2008) Development of a method for the quantification of whey allergen traces in mixed-fruit juices based on liquid chromatography with mass spectrometric detection. *J. Chromatogr. A* **1192:** 113–120

Mota I, Piedade S, Gaspar Â, Benito-Garcia F, Sampaio G, Borrego LM & Morais-Almeida M (2018) Cow's milk oral immunotherapy in real life: 8-year long-term follow-up study. *Asia Pac. Allergy* **8:** e28

Muraro A, Hoffmann-Sommergruber K, Holzhauser T, Poulsen LK, Gowland MH, Akdis CA, Mills ENC, Papadopoulos N, Roberts G, Schnadt S, Van Ree R, Sheikh A & Vieths S (2014a) EAACI Food Allergy and Anaphylaxis Guidelines. Protecting consumers with food allergies: Understanding food consumption, meeting regulations and identifying unmet needs. *Allergy Eur. J. Allergy Clin. Immunol.* **69:** 1464–1472

Muraro A, Werfel T, Hoffmann-Sommergruber K, Roberts G, Beyer K, Bindslev-Jensen C, Cardona V, Dubois A, DuToit G, Eigenmann P, Fernandez Rivas M, Halken S, Hickstein L, Høst A, Knol E, Lack G, Marchisotto MJ, Niggemann B, Nwaru BI, Papadopoulos NG, et al (2014b) EAACI food allergy and anaphylaxis guidelines: diagnosis and management of food allergy. *Allergy* **69:** 1008–1025

Navuluri L, Parvataneni S, Hassan H, Birmingham NP, Kelly C & Gangur V (2006) Allergic and anaphylactic response to sesame seeds in mice: Identification of Ses i 3 and basic subunit of 11s globulins as allergens. *Int. Arch. Allergy Immunol.* **140:** 270–276

Newsome GA & Scholl PF (2013) Quantification of allergenic bovine milk αs1-casein in baked goods using an intact 15N-labeled protein internal standard. *J. Agric. Food Chem.* **61:** 5659–5668

Nielen MM (2017) Smartphone analyzers for on-site testing of food quality and safety-FoodSmartphone.

Nollet LML & van Hengel AJ (2010) Food Allergens: Analysis Instrumentation and Methods CRC Press

Nowatzke WL, Oliver KG, Cho CY, Rallabhandi P & Garber EAE (2019) Single-Laboratory Validation of the Multiplex xMAP Food Allergen Detection Assay with Incurred Food Samples. *J. Agric. Food Chem.* **67:** 484–498

NVWA (2016) Advice on preliminary reference doses for food allergens Available at: https://english.nvwa.nl

Nwaru BI, Hickstein L, Panesar SS, Muraro A, Werfel T, Cardona V, Dubois AEJ, Halken S, Hoffmann-Sommergruber K, Poulsen LK, Roberts G, Van Ree R, Vlieg-Boerstra BJ & Sheikh A (2014) The epidemiology of food allergy in Europe: A systematic review and meta-analysis. *Allergy Eur. J. Allergy Clin. Immunol.* **69:** 62–75

Olcese R, Silvestri M, Del Barba P, Brolatti N, Barberi S, Tosca MA & Ciprandi G (2019) Mal d 1 and Bet v 1 sensitization pattern in children with Pollen Food Syndrome. *Allergol. Int.* **68:** 122–124

de Oliveira FC, Coimbra JS dos R, de Oliveira EB, Zuñiga ADG & Rojas EEG (2016) Food Protein-polysaccharide Conjugates Obtained via the Maillard Reaction: A Review. *Crit. Rev. Food Sci. Nutr.* **56:** 1108–1125

Olsen J V., Ong SE & Mann M (2004) Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics* **3:** 608–614

Pace CN, Vajdos F, Fee L, Grimsley G & Gray T (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4:** 2411–2423

Paez V, Barrett WB, Deng X, Diaz-Amigo C, Fiedler K, Fuerer C, Hostetler GL, Johnson P, Joseph G, Konings EJM, Lacorn M, Lawry J, Liu H, Marceau E, Mastovska K, Monteroso L, Pan SJ, Parker C, Phillips MM, Popping B, et al (2016) AOAC SMPR(®) 2016.002. *J. AOAC Int.* **99:** 1122–1124

Palladino C (2018) Peanut allergens. *Mol. Immunol.* **100:** 58–70

Di Palma S, Hennrich ML, Heck AJR & Mohammed S (2012) Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. *J. Proteomics* **75:** 3791–3813

Pauwels J, Lamberty A & Schimmel H (1998) Homogeneity testing of reference materials. *Accredit. Qual. Assur.* **3:** 51–55

Perutka Z & Sebela M (2018) Pseudotrypsin: A little-known trypsin proteoform. *Molecules* **23:** 2637

Peters RL, Allen KJ, Dharmage SC, Koplin JJ, Dang T, Tilbrook KP, Lowe A, Tang MLK & Gurrin LC (2015) Natural history of peanut allergy and predictors of resolution in the first 4 years of life: A population-based assessment. *J. Allergy Clin. Immunol.* **135:** 1257–1266

Picó Y, Monaci L & Visconti A (2012) Allergens. In *Chemical Analysis of Food: Techniques and Applications* pp 693–714. Academic Press

Pilolli R, De Angelis E & Monaci L (2020) 19 - Advances in MS methods for food allergens detection. In, Pico YBT-CA of F (Second E (ed) pp 787–811. Academic Press Available at: http://www.sciencedirect.com/science/article/pii/B978012813266100019X

Planque M, Arnould T, Delahaut P, Renard P, Dieu M & Gillard N (2019) Development of a strategy for the quantification of food allergens in several food products by mass spectrometry in a routine laboratory. *Food Chem.* **274:** 35–45

Planque M, Arnould T, Dieu M, Delahaut P, Renard P & Gillard N (2016) Advances in ultra-high performance liquid chromatography coupled to tandem mass spectrometry for sensitive detection of several food allergens in complex and processed foodstuffs. *J. Chromatogr. A* **1464:** 115–123

Planque M, Arnould T & Gillard N (2017a) Food Allergen Analysis: Detection, Quantification and Validation by Mass Spectrometry. In *Allergen* p 7.

Planque M, Arnould T, Renard P, Delahaut P, Dieu M & Gillard N (2017b) Highlight on bottlenecks in food allergen analysis: Detection and quantification by mass spectrometry. *J. AOAC Int.* **100:** 1126–1130

Platteau C, De Loose M, De Meulenaer B & Taverniers I (2011) Quantitative detection of hazelnut (Corylus avellana) in cookies: ELISA versus real-time PCR. *J. Agric. Food Chem.* **59:** 11395–11402

Pomés A, Davies JM, Gadermaier G, Hilger C, Holzhauser T, Lidholm J, Lopata AL, Mueller GA, Nandy A, Radauer C, Chan SK, Jappe U, Kleine-Tebbe J, Thomas WR, Chapman MD, van Hage M, van

Ree R, Vieths S, Raulf M & Goodman RE (2018) WHO/IUIS Allergen Nomenclature: Providing a common language. *Mol. Immunol.* **100:** 3–13

Poms RE, Klein CL & Anklam E (2004) Methods for allergen analysis in food: A review. *Food Addit. Contam.* **21:** 1–31

Popping B & Diaz-Amigo C (2018) European regulations for labeling requirements for food allergens and substances causing intolerances: History and future. *J. AOAC Int.* **101:** 2–7

Posada-Ayala M, Alvarez-Llamas G, Maroto AS, Maes X, Muñoz-Garcia E, Villalba M, Rodríguez R, Perez-Gordo M, Vivanco F, Pastor-Vargas C & Cuesta-Herranz J (2015) Novel liquid chromatography-mass spectrometry method for sensitive determination of the mustard allergen Sin a 1 in food. *Food Chem.* **183:** 58–63

Poulos LM, Waters AM, Correll PK, Loblay RH & Marks GB (2007) Trends in hospitalizations for anaphylaxis, angioedema, and urticaria in Australia, 1993-1994 to 2004-2005. *J. Allergy Clin. Immunol.* **120:** 878–884

Pratt JM, Simpson DM, Doherty MK, Rivers J, Gaskell SJ & Beynon RJ (2006) Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nat. Protoc.* **1:** 1029

Prescott S & Allen KJ (2011) Food allergy: Riding the second wave of the allergy epidemic. *Pediatr. Allergy Immunol.* **22:** 155–160

Raijmakers R, Neerincx P, Mohammed S & Heck AJR (2010) Cleavage specificities of the brother and sister proteases Lys-C and Lys-N. *Chem. Commun.* **46:** 8827–8829

Ramsdell F & Fowlkes BJ (1990) Clonal deletion versus clonal anergy: The role of the thymus in inducing self tolerance. *Science (80-. ).* **248:** 1342–1348

Randall KL & Hawkins CA (2018) Antihistamines and allergy. *Aust. Prescr.* **41:** 41

Remington BC (2013) Risk Assessment of Trace and Undeclared Allergens in Processed Foods. *Diss. Theses Food Sci. Technol.*

Remington BC, Westerhout J, Meima MY, Blom WM, Kruizinga AG, Wheeler MW, Taylor SL, Houben GF & Baumert JL (2020) Updated population minimal eliciting dose distributions for use in risk assessment of 14 priority food allergens. *Food Chem. Toxicol.* **139:**

Renz H, Allen KJ, Sicherer SH, Sampson HA, Lack G, Beyer K & Oettgen HC (2018) Food allergy. *Nat. Rev. Dis. Prim.* **4:** 1–20

Rice RH, Means GE & Brown WD (1977) Stabilization of bovine trypsin by reductive methylation. *Biochim. Biophys. Acta (BBA)-Protein Struct.* **492:** 316–321

Röder M, Vieths S & Holzhauser T (2011) Sensitive and specific detection of potentially allergenic almond (Prunus dulcis) in complex food matrices by Taqman® real-time polymerase chain reaction in comparison to commercially available protein-based enzyme-linked immunosorbent assay. *Anal. Chim. Acta* **685:** 74–83

Rodriguez J, Gupta N, Smith RD & Pevzner PA (2008) Does trypsin cut before proline? *J. Proteome Res.* **7:** 300–305

Rodríguez-González P & Ignacio García Alonso J (2019) Mass Spectrometry | Isotope dilution mass spectrometry. In *Encyclopedia of Analytical Science* pp 411–420. Elsevier

Romano A & Castagna V (2016) Diet as Therapy for IBD? In *Human Nutrition From the Gastroenterologist's Perspective: Lessons From Expo Milano 2015* pp 79–98.

Rona RJ, Keil T, Summers C, Gislason D, Zuidmeer L, Sodergren E, Sigurdardottir ST, Lindner T, Goldhahn K, Dahlstrom J, McBride D & Madsen C (2007) The prevalence of food allergy: A meta-analysis. *J. Allergy Clin. Immunol.* **120:** 638–646

Ross GMS, Bremer MGEG & Nielen MWF (2018) Consumer-friendly food allergen detection: moving towards smartphone-based immunoassays. *Anal. Bioanal. Chem.* **410:** 5353–5371

Rudolf J, Ansari P, Kern C, Ludwig T & Baumgartner S (2012) Effects of different extraction buffers on peanut protein detectability and lateral flow device (LFD) performance. *Food Addit. Contam. - Part A Chem. Anal. Control. Expo. Risk Assess.* **29:** 1–11

Russell MR, Achour B, McKenzie EA, Lopez R, Harwood MD, Rostami-Hodjegan A & Barber J (2013) Alternative fusion protein strategies to express recalcitrant QconCAT proteins for quantitative proteomics of human drug metabolizing enzymes and transporters. *J. Proteome Res.* **12:** 5934–5942

Sampson MA, Muñoz-Furlong A & Sicherer SH (2006) Risk-taking and coping strategies of adolescents and young adults with food allergy. *J. Allergy Clin. Immunol.* **117:** 1440–1445

Sanchez-Garcia L, Martín L, Mangues R, Ferrer-Miralles N, Vázquez E & Villaverde A (2016) Recombinant pharmaceuticals from microbial cells: a 2015 update. *Microb. Cell Fact.* **15:** 33

Savage J & Johns CB (2015) Food Allergy: Epidemiology and Natural History. *Immunol. Allergy Clin. North Am.* **35:** 45–49

Scala E, Abeni D, Cecchi L, Guerra EC, Locanto M, Pirrotta L, Giani M & Asero R (2017) Molecular Recognition Profiles and Clinical Patterns of PR-10 Sensitization in a Birch-Free Mediterranean Area. *Int. Arch. Allergy Immunol.* **173:** 138–146

Schein CH & Noteborn MHM (1988) Formation of soluble recombinant proteins in escherichia coli is favored by lower growth temperature. *Bio/Technology* **6:** 291–294

Schmidt A, Picotti P & Aebersold R (2008) Systembiologie Proteomanalyse und Systembiologie. *BioSpektrum* **14:** 44

Schneider N, Becker CM & Pischetsrieder M (2010) Analysis of lysozyme in cheese by immunocapture mass spectrometry. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **878:** 201–206

Schoemaker AA, Sprikkelman AB, Grimshaw KE, Roberts G, Grabenhenrich L, Rosenfeld L, Siegert S, Dubakiene R, Rudzeviciene O, Reche M, Fiandor A, Papadopoulos NG, Malamitsi-Puchner A, Fiocchi A, Dahdah L, Sigurdardottir ST, Clausen M, Stańczyk-Przyłuska A, Zeman K, Mills ENC, et al (2015) Incidence and natural history of challenge-proven cow's milk allergy in European children - EuroPrevall birth cohort. *Allergy Eur. J. Allergy Clin. Immunol.* **70:** 963–972

Schrader C, Schielke A, Ellerbroek L & Johne R (2012) PCR inhibitors - occurrence, properties and removal. *J. Appl. Microbiol.* **113:** 1014–1026

SciCom (2017) AVIS 24-2017: Doses de référence pour les allergènes repris à l'annexe II du Règlement (UE) n° 1169/2011 du 25 octobre 2011. FAFSC. Available at: http://www.afsca.be/comitescientifique/avis/2017/_documents/Avis24-2017_SciCom2017-01_dosesdereferenceallergenes.pdf

Searle BC (2010) Scaffold: A bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics* **10:** 1265–1269

Senyuva HZ, Jones IB, Sykes M & Baumgartner S (2019) A critical review of the specifications and performance of antibody and DNA-based methods for detection and quantification of allergens in foods. *Food Addit. Contam. - Part A Chem. Anal. Control. Expo. Risk Assess.* **36:** 507–547

Shabalina SA, Ogurtsov AY & Spiridonov NA (2006) A periodic pattern of mRNA secondary structure created by the genetic code. *Nucleic Acids Res.* **34:** 2428–2437

Sharma GM, Khuda SE, Parker CH, Eischeid AC & Pereira M (2017) Detection of Allergen Markers in Food: Analytical Methods. In *Food Safety: Innovative Analytical Tools for Safety Assessment*, Spizzirri UG & Cirillo G (eds) pp 65–121.

Shenoy P, Viau M, Tammel K, Innings F, Fitzpatrick J & Ahrné L (2015) Effect of powder densities, particle size and shape on mixture quality of binary food powder mixtures. *Powder Technol.* **272:** 165–172

Sicherer SH & Sampson HA (2006) 9. Food allergy. *J. Allergy Clin. Immunol.* **117:** S470–S475

Sicherer SH & Sampson HA (2010) Food allergy. *J. Allergy Clin. Immunol.* **125:** S116–S125

Sicherer SH & Sampson HA (2014) Food allergy: Epidemiology, pathogenesis, diagnosis, and treatment. *J. Allergy Clin. Immunol.* **133:** 291–307

Siepen JA, Keevil EJ, Knight D & Hubbard SJ (2007) Prediction of missed cleavage sites in tryptic peptides aids protein identification in proteomics. *J. Proteome Res.* **6:** 399–408

De Silva D, Geromi M, Panesar SS, Muraro A, Werfel T, Hoffmann-Sommergruber K, Roberts G, Cardona V, Dubois AEJ, Halken S, Host A, Poulsen LK, Van Ree R, Vlieg-Boerstra BJ, Agache I & Sheikh A (2014) Acute and long-term management of food allergy: Systematic review. *Allergy Eur. J. Allergy Clin. Immunol.* **69:** 159–167

Słowianek M & Majak I (2011) Biotechnology and Food Science: Methods of allergen detection based on DNA analysis. *Biotechnol Food Sci* **75:** 39–44

Sriram P, Kalogerakis N & Behie LA (1996) Experimental determination of the rate of autolysis of trypsin at 37°C. *Biotechnol. Tech.* **10:** 601–606

Strid J, Thomson M, Hourihane J, Kimber I & Strobel S (2004) A novel model of sensitization and oral tolerance to peanut protein. *Immunology* **113:** 293–303

Suh SM, Park SB, Kim MJ & Kim HY (2019) Simultaneous detection of fruit allergen-coding genes in tomato, apple, peach and kiwi through multiplex PCR. *Food Sci. Biotechnol.* **28:** 1593–1598

Tordesillas L, Berin MC & Sampson HA (2017) Immunology of Food Allergy. *Immunity* **47:** 32–50

Triggiani M, Patella V, Staiano RI, Granata F & Marone G (2008) Allergy and the cardiovascular system. *Clin. Exp. Immunol.* **153:** 7–11

Turner PJ, Gowland MH, Sharma V, Ierodiakonou D, Harper N, Garcez T, Pumphrey R & Boyle RJ (2015) Increase in anaphylaxis-related hospitalizations but no increase in fatalities: An analysis of United Kingdom national anaphylaxis data, 1992-2012. *J. Allergy Clin. Immunol.* **135:** 956–963

Valenta R, Hochwallner H, Linhart B & Pahr S (2015) Food allergies: The basics. *Gastroenterology* **148:** 1120–1131

Vickery BP, Chin S & Burks AW (2011) Pathophysiology of Food Allergy. *Pediatr. Clin. North Am.* **58:** 363–376

Vighi G, Marcucci F, Sensi L, Di Cara G & Frati F (2008) Allergy and the gastrointestinal system. *Clin. Exp. Immunol.* **153:** 3–6

Villa C, Costa J, Gondar C, Oliveira MBPP & Mafra I (2018) Effect of food matrix and thermal processing on the performance of a normalised quantitative real-time PCR approach for lupine (Lupinus albus) detection as a potential allergenic food. *Food Chem.* **262:** 251–259

Villanueva J, Carrascal M & Abian J (2014) Isotope dilution mass spectrometry for absolute quantification in proteomics: Concepts and strategies. *J. Proteomics* **96:** 184–199

Van Vlierberghe K, Gavage M, Dieu M, Renard P, Arnould T, Gillard N, Coudijzer K, De Loose M, Gevaert K & Van Poucke C (2020) Selection of universal peptide biomarkers for the detection of the allergen hazelnut in food trough a comprehensive, high resolution mass spectrometric (HRMS) based approach. *Food Chem.* **309:** 125679

Walczyk NE, Smith PMC, Tovey E, Wright GC, Fleischfresser DB & Roberts TH (2013) Analysis of crude protein and allergen abundance in peanuts (Arachis hypogaea cv. Walter) from three growing regions in Australia. *J. Agric. Food Chem.* **61:** 3714–3725

Waldo GS, Standish BM, Berendzen J & Terwilliger TC (1999) Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* **17:** 691–695

Walker MJ (2018) Food allergens: An update on analytical methods. In *Encyclopedia of Food Chemistry*, Melton LD Shahidi F & Varelis P (eds) pp 622–639. Elsevier Inc.

Walker MJ, Burns DT, Elliott CT, Gowland MH & Mills ENC (2016) Is food allergen analysis flawed? Health and supply chain risks and a proposed framework to address urgent analytical needs. *Analyst* **141:** 24–35

Wang J, Vanga SK, McCusker C & Raghavan V (2019) A Comprehensive Review on Kiwifruit Allergy: Pathogenesis, Diagnosis, Management, and Potential Modification of Allergens Through Processing. *Compr. Rev. Food Sci. Food Saf.* **18:** 500–513

Wang WQ, Jensen ON, Møller IM, Hebelstrup KH & Rogowska-Wrzesinska A (2018) Evaluation of sample preparation methods for mass spectrometry-based proteomic analysis of barley leaves. *Plant Methods* **14:** 72

Wu Z, Zhou N, Xiong F, Li X, Yang A, Tong P, Tang R & Chen H (2016) Allergen composition analysis and allergenicity assessment of Chinese peanut cultivars. *Food Chem.* **196:** 459–465

Yu W, Freeland DMH & Nadeau KC (2016) Food allergy: Immune mechanisms, diagnosis and immunotherapy. *Nat. Rev. Immunol.* **16:** 751

Zhang H, Lu Y, Ushio H & Shiomi K (2014a) Development of sandwich ELISA for detection and quantification of invertebrate major allergen tropomyosin by a monoclonal antibody. *Food Chem.* **150:** 151–157

Zhang J, Lai S, Cai Z, Chen Q, Huang B & Ren Y (2014b) Determination of bovine lactoferrin in dairy products by ultra-high performance liquid chromatography-tandem mass spectrometry based on tryptic signature peptides employing an isotope-labeled winged peptide as internal standard. *Anal. Chim. Acta* **829:** 33–39

Zhang J, Lai S, Zhang Y, Huang B, Li D & Ren Y (2012) Multiple reaction monitoring-based determination of bovine α-lactalbumin in infant formulas and whey protein concentrates by ultra-high performance liquid chromatography-tandem mass spectrometry using tryptic signature peptides and synthetic peptide stand. *Anal. Chim. Acta* **727:** 47–53

**ANNEX 1: UNamur in-gel protein digestion protocol for HRMS analysis**

!! Wearing disposable nitrile powder free gloves is mandatory for all the following steps !!

Work cleanly and cautiously to reduce keratin and chemical background contaminations of the samples. Use an agitator for Eppendorf tubes which allows rate 300-1200 rpm and temperature settings.

A. Sample preparation

1. Excise the bands of interest (blue spots) from the gel using a scalpel.
2. Cut the resulting excised spots into small cubes (ca. 1 mm). Note that smaller pieces could clog pipette tips.
3. Transfer gel pieces from the corresponding bands into a Protein LoBind microcentrifuge tube of 1.5 mL.

These samples are placed in ULC/MS grade water and stocked at -20°C up to digestion and analysis. Be sure that all the gel pieces are completely covered with water.

NB: the volumes mentioned in the following procedure (steps B to E) can be adjusted to completely cover the gel pieces.

B. Destaining of the gel pieces excised from Coomassie-stained gels

1. Incubate the samples for 10 min (21 °C, 900 rpm)
2. Remove the solvent, while ensuring that the gel pieces are not eliminated
3. Add 50 µL of ammonium bicarbonate buffer 50 mM (ABB 50), and incubate for 5 min (21°C; 900 rpm)
4. Remove all liquid.
5. Add 50 µL of ACN, incubate for 5 min (21°C; 900 rpm), and then remove all liquid to destain the gel pieces.
6. Repeat steps 3 to 5 twice or three times to completely destain the gel pieces.

C. In-gel reduction and alkylation of the protein cysteines

1. Add 40 µL of freshly prepared DTT 10 mM
2. Incubate the sample at 56 °C for 45 min (600 rpm)
3. Cool the microcentrifuge tubes to room temperature, remove the DTT solution
4. Add 40 µL of freshly prepared IAA 55 mM
5. Incubate at 21°C in the darkness for 30 min (600 rpm)
6. Remove all liquid
7. Add 50 µL of washing solution (water/ACN 1/1 (v/v)) to the gel pieces, and agitate for 5 min (21°C; 900 rpm)
8. Remove all liquid
9. Add 50 µL of ACN, and agitate for 10 min (21°C; 900 rpm)
10. Remove all liquid
11. Add 50 µL of ABB 100, and agitate for 5 min (21°C; 900 rpm)
12. Remove all liquid
13. Add 50 µL of ACN, and agitate for 10 min (21°C; 900 rpm)
14. Remove all liquid
15. Dry the gel pieces for 15 min at 40°C (microcentrifuge tubes open)

NB: steps 11 to 14 can be repeated a second times if needed.

D. In-gel digestion of the protein with trypsin

1. Place all the microcentrifuge tubes in ice bath for at least 15 min before to perform the following steps
2. Dilute 4 times the iced trypsin solution (0.1 µg/µL into iced acetic acid 50 mM; defrosted just before use) with ABB 50
3. Add 30 µL of diluted iced trypsin solution (diluted 4 times with ABB 50) to cover all the gel pieces (more trypsin can be added if this is required to recover the initial size of the gel)
4. Incubate for 30 to 45 min on ice bath
5. Remove the excess of trypsin
6. Add 50 µL of ABB 50
7. Incubate overnight at 37°C (300 rpm)

E. Extract recovery

1. Transfer the digested solution into a new microcentrifuge tube
2. Add 50 µL of ACN to the gel pieces
3. Incubate for 30 min at 37 °C (900 rpm)
4. Transfer all liquid to the microcentrifuge tube containing tryptic digest solution
5. Evaporate ACN for 60 min at 40°C
6. The resulting solution is transferred into UPLC plastic vial and can be store at -20°C up to its analysis.

**ANNEX 2: DNA sequences of the different designed constructs for concatemer production**

Construct 1-1: Construct not coding for any peptide biomarker that contains coding sequences for GST primer, GFP and poly-H tag with an expected product of 30 kDa.

CATATGTCTCCTATACTAGGTTATCCCGGGGGATCCCCATGGGGTCTAGAACCCGGGGATATCATGGCTAGCAAAGGAGAAGAACTTTTCACTG
GAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCTACATACGGAAA
GCTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCGTGGCCAACACTTGTCACTACTTTCTCTTATGGTGTTCAATGCTTTTCC
CGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACG
GGAACTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAA
CATTCTCGGACACAAACTGGAGTACAACTATAACTCACACAATGTATACATCACGGCAGACAAACAAAAGAATGGAATCAAAGCTAACTTCAAA
ATTCGCCACAACATTGAAGATGGCTCCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACA
ACCATTACCTGTCGACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGAT
TACACATGGCATGGATGAGCTCTACAAAGATATCCCGGCGAGCCATCACCATCACCATCACCATTAATGACTCGAG

Construct 1-2: Construct coding for a GST primer, concatenated egg peptides, GFP and poly-H tag with an expected product of 41 kDa. One copy of each one of the 8 egg peptide biomarkers are placed in this order: EW1-EW2-EW4-EW3-EY2-EY1-EY3-EY4.

CATATGTCTCCTATACTAGGTTATCCCGGGGGATCCCGCGGCGGCCTGGAACCGATTAACTTTCAGACCGCGGCGGATCAGGCGCGTCATATTG
CGACCAACGCGGTGCTGTTTTTTGGCCGTTTTGAAAGCAACTTTAACACCCAGGCGACCAACCGTAGCGCGGGCTGGAACATTCCGATTGGCAC
CCTGATTCATCGTAACGTGAACTTTGATGGCGAAATTCTGAAAGCGACCGCGGTGAGCCTGCTGGAATGGCAGCGTAACTTTCTGATTAACGAA
ACCGCGCGTCATCTGTTTCTGCCGAGCAGCTATAAAGGGGGATCCCCATGGGGTCTAGAACCCGGGGATATCATGGCTAGCAAAGGAGAAGAAC
TTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCTAC
ATACGGAAAGCTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCGTGGCCAACACTTGTCACTACTTTCTCTTATGGTGTTCAA
TGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCA
AAGATGACGGGAACTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAAGGTATTGATTTTAAAGA
AGATGGAAACATTCTCGGACACAAACTGGAGTACAACTATAACTCACACAATGTATACATCACGGCAGACAAACAAAAGAATGGAATCAAAGCT
AACTTCAAAATTCGCCACAACATTGAAGATGGCTCCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTT
TACCAGACAACCATTACCTGTCGACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGC
TGCTGGGATTACACATGGCATGGATGAGCTCTACAAAGATATCCCGGCGAGCCATCACCATCACCATCACCATTAATGACTCGAG

Construct 1-3: Construct coding for a GST primer, concatenated egg peptides, GFP and poly-H tag with an expected product of 59 kDa. Only egg white peptide biomarkers are considered in this construct. DNA sequences corresponding to peptides from abundant egg white protein are repeated to obtain egg white peptides in natural abundance. DNA sequences for concatenated egg white peptides followed this sequence: EW1-EW2-EW4-EW1-EW2-EW4-EW1-EW2-EW3-EW1-EW2-EW1-EW2-EW1-EW2-EW1-EW2.

CATATGTCTCCTATACTAGGTTATCCCGGGGGATCCCCATGGCGCGGTGGCCTCGAACCAATTAACTTTCAGACGGCGGCCGATCAGGCCCGTC
ATATTGCCACCAACGCCGTGCTCTTTTTTGGTCGTTTTGAAAGCAACTTTAACACCCAGGCGACCAACCGTGGCGGCCTGGAACCGATTAACTT
TCAGACCGCGGCGGATCAGGCGCGTCATATTGCGACCAACGCGGTGCTGTTTTTTGGCCGTTTCGAGAGTAATTTCAATACGCAAGCCACGAAT
CGCGGCGGTCTGGAGCCGATTAATTTTCAGACCGCCGCGGATCAAGCGCGTCACATTGCGACGAACGCGGTCCTGTTCTTTGGCCGTAGCGCGG
GCTGGAACATTCCGATTGGCACCCTGATTCATCGTGGCGGGCTGGAACCAATTAACTTCCAGACCGCGGCAGATCAGGCACGTCATATTGCAAC
CAACGCAGTGCTTTTTTTTGGGCGTGGGGGCCTTGAACCTATTAACTTTCAAACCGCGGCAGATCAAGCGCGTCACATTGCGACTAACGCGGTC
CTGTTTTTTGGCCGCGGCGGTCTCGAACCAATTAACTTTCAAACCGCGGCGGACCAAGCGCGTCATATCGCCACCAACGCGGTTCTCTTTTTTG
GTCGTGGCGGTCTGGAACCAATCAACTTTCAAACGGCGGCGGACCAAGCGCGTCACATCGCGACCAACGCCGTTCTGTTTTTTGGTCGCGGCGG
CCTTGAGCCGATTAACTTCCAAACCGCAGCAGATCAGGCGCGCCATATCGCAACCAACGCAGTCCTGTTTTTCGGTCGTGGGCCATGGGGTCTA
GAACCCGGGGATATCATGGCTAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACA
AATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCTACATACGGAAAGCTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCGTG
GCCAACACTTGTCACTACTTTCTCTTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGTGCCATG
CCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGGAACTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTG
TTAATCGTATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTCGGACACAAACTGGAGTACAACTATAACTCACACAATGTATA
CATCACGGCAGACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACATTGAAGATGGCTCCGTTCAACTAGCAGACCATTAT
CAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGCCCTTTCGAAAGATCCCAACGAAA
AGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAAGATATCCCGGCGAGCCATCA
CCATCACCATCACCATTAATGACTCGAG

Construct 1-4: Construct coding for a GST primer, concatenated egg peptides, GFP and poly-H with an expected product of 42 kDa. Only egg yolk peptide biomarkers are considered in this construct. Peptides from abundant egg yolk protein are repeated to obtain egg yolk peptides in natural abundance. Concatenated egg yolk peptides followed this sequence: EY3-EY2-EY1-EY2-EY1-EY3-EY3-EY4-EY3-EY3

CATATGTCTCCTATACTAGGTTATCCCGGGGGATCCCCATGGGGTCTAGAACGTCATCTGTTTCTGCCGAGCAGCTATAAAAACGTGAACTTTG
ATGGCGAAATTCTGAAAGCGACCGCGGTGAGCCTGCTGGAATGGCAGCGTAACGTTAACTTTGATGGTGAAATTCTCAAAGCCACCGCCGTGAG
CCTCCTGGAATGGCAGCGCCATCTCTTTCTGCCAAGCAGCTATAAACATCTGTTCCTCCCGAGCAGCTATAAAAACTTTCTGATTAACGAAACC
GCGCGTCATCTCTTTCTCCCGAGTAGCTATAAACATCTGTTCCTGCCAAGCAGTTATAAAGGTCTAGAACCCGGGGATATCATGGCTAGCAAAG
GAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGG
TGATGCTACATACGGAAAGCTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCGTGGCCAACACTTGTCACTACTTTCTCTTAT
GGTGTTCAATGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTA
TATCTTTCAAAGATGACGGGAACTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAAGGTATTGA
TTTTAAAGAAGATGGAAACATTCTCGGACACAAACTGGAGTACAACTATAACTCACACAATGTATACATCACGGCAGACAAACAAAAGAATGGA
ATCAAAGCTAACTTCAAAATTCGCCACAACATTGAAGATGGCTCCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCC
CTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTT
TGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAAGATATCCCGGCGAGCCATCACCATCACCATCACCATTAATGACTCGAG

Construct 2-1: Construct with optimized mRNA secondary structure coding for a GST primer, concatenated egg peptides, GFP and poly-H with an expected product 41 kDa. One copy of each one of the 8 egg peptide biomarkers are placed in this order: EW3-EW1-EW2-EW4-EY1-EY4-EY3-EY2.

CATATGTCTCCTATTCTAGGTTATCCAGGTGGTTCGCGCTCTGCAGGTTGGAACATTCCGATTGGAACGCTGATTCATCGTGGCGGCCTGGAAC
CGATTAACTTTCAAACCGCGGCGGATCAGGCGCGTCACATTGCCACCAATGCCGTCCTGTTTTTTGGCCGTTTTGAAAGCAACTTTAACACCCA
GGCGACTAACCGTGCTACGGCTGTATCTCTCCTGGAATGGCAGCGTAACTTTCTGATTAACGAAACCGCGCGTCATCTGTTTCTGCCGAGCAGC
TATAAAAACGTCAACTTTGATGGCGAAATTCTGAAGGGGGGATCTCCGTGGGGCCTAGAACCAGGAGATATCATGGCTAGCAAAGGAGAAGAAC
TTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCTAC
ATACGGCAAGCTAACCCTGAAGTTTATCTGTACAACTGGCAAGTTACCGGTACCCTGGCCAACACTTGTCACTACTTTCTCGTATGGTGTACAA
TGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTTAAAAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACCATATCTTTCA
AAGATGACGGCAACTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAAGGTATTGATTTCAAAGA
AGATGGTAACATTCTCGGCCACAAACTGGAGTACAACTATAACTCACACAATGTATACATTACGGCGGATAAACAGAAGAATGGAATCAAGGCT
AACTTCAAAATTCGCCACAACATTGAAGATGGCTCCGTTCAGTTGGCAGATCATTACCAACAGAATACCCCAATTGGCGATGGCCCTGTCCTTT
TACCAGACAACCATTACCTGTCGACACAGAGCGCTCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGC
TGCTGGGATTACACATGGCATGGATGAGCTCTACAAAGATATCCCGGCGAGCCATCACCATCACCATCACCATTAACTCGAG

Construct 2-2: Construct with optimized mRNA secondary structure coding for a GST primer, concatenated egg peptides, GFP and poly-H with an expected product 41 kDa. This construct is similar to construct 2-1, with N-term methionine codon ATG of GFP sequence replaced by leucine codon TTA. One copy of each one of the 8 egg peptide biomarkers are placed in this order: EW3-EW1-EW2-EW4-EY1-EY4-EY3-EY2.

CATATGTCTCCTATTCTAGGTTATCCAGGTGGTTCGCGCTCTGCAGGTTGGAACATTCCGATTGGAACGCTGATTCATCGTGGCGGCCTGGAAC
CGATTAACTTTCAAACCGCGGCGGATCAGGCGCGTCACATTGCCACCAATGCCGTCCTGTTTTTTGGCCGTTTTGAAAGCAACTTTAACACCCA
GGCGACTAACCGTGCTACGGCTGTATCTCTCCTGGAATGGCAGCGTAACTTTCTGATTAACGAAACCGCGCGTCATCTGTTTCTGCCGAGCAGC
TATAAAAACGTCAACTTTGATGGCGAAATTCTGAAGGGGGGATCTCCGTGGGGCCTAGAACCAGGAGATATCTTAGCTAGCAAAGGAGAAGAAC
TTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCTAC
ATACGGCAAGCTAACCCTGAAGTTTATCTGTACAACTGGCAAGTTACCGGTACCCTGGCCAACACTTGTCACTACTTTCTCGTATGGTGTACAA
TGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTTAAAAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACCATATCTTTCA
AAGATGACGGCAACTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAAGGTATTGATTTCAAAGA
AGATGGTAACATTCTCGGCCACAAACTGGAGTACAACTATAACTCACACAATGTATACATTACGGCGGATAAACAGAAGAATGGAATCAAGGCT
AACTTCAAAATTCGCCACAACATTGAAGATGGCTCCGTTCAGTTGGCAGATCATTACCAACAGAATACCCCAATTGGCGATGGCCCTGTCCTTT
TACCAGACAACCATTACCTGTCGACACAGAGCGCTCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGC
TGCTGGGATTACACATGGCATGGATGAGCTCTACAAAGATATCCCGGCGAGCCATCACCATCACCATCACCATTAACTCGAG

Construct 2-3: Construct with optimized mRNA secondary structure coding for a GST primer, concatenated egg peptides and poly-H with an expected product 14 kDa. One copy of each one of the 8 egg peptide biomarkers are placed in this order: EW3-EW1-EW2-EW4-EY1-EY4-EY3-EY2. This construct is similar to construct 2-1 with GFP sequence deletion.

CATATGTCTCCTATTCTAGGTTATCCAGGTGGTTCGCGCTCTGCAGGTTGGAACATTCCGATTGGAACGCTGATTCATCGTGGCGGCCTGGAAC
CGATTAACTTTCAAACCGCGGCGGATCAGGCGCGTCACATTGCCACCAATGCCGTCCTGTTTTTTGGCCGTTTTGAAAGCAACTTTAACACCCA
GGCGACTAACCGTGCTACGGCTGTATCTCTCCTGGAATGGCAGCGTAACTTTCTGATTAACGAAACCGCGCGTCATCTGTTTCTGCCGAGCAGC
TATAAAAAACGTCAACTTTGATGGCGAAATTCTGAAGGGGGGATCTCCGTGGGGCCTAGAACCAGGAGATATCCCGGCGAGCCATCACCATCACC
ATCACCATTAACTCGAG

Construct 3-1: Construct with optimized mRNA secondary structure coding for a GST primer, concatenated 19 final peptide biomarkers and poly-H with an expected product 29 kDa. One copy of each one of the 19 final peptide biomarkers are placed in this order: P13-P4-P1-P5-P2-P8-P16-P18-P3-P7-P12-P14-P11-P10-P6-P17-P9-P15-P19

CATATGTCTCCTATTCTAGGTTACCCAGGTGGTTCACGTGGCTCCGAGGAAGAAGATATTACCAACCCTATCAACTTGCGCATTGACGCACTGA
ACGAAAACAAATTTTTTGTGGCGCCGTTTCCGGAAGTGTTCGGCAAAACCCCCGAAGTGGATGATGAAGCGCTGGAAAAACATCAGGGCCTGCC
GCAAGAAGTGCTGAACGAAAACCTGCTGCGCCATATTGCGACCAACGCAGTGTTGTTTTTTGGCCGCGCAGATATTTATACCGAACAGGTAGGC
CGTCTGAACGCACTGGAGCCAACCAACCGTTATCTGGGCTACCTGGAACAACTGCTGCGGGGCGGCCTGGAACCGATTAACTTTCAGACCGCAG
CAGATCAGGCGCGCACCGTAATTGTCGAAGCTCCAATTCATGGCCTGAAAGGCAGCGAAGAAGAAGGCGACATTACCAACCCGATTAACCTGCG
TAACGTTAACTTTGATGGCGAAATTCTGAAAAGCGCGGGCTGGAACATTCCGATTGGCACCCTGATTCATCGCGTCTATGTGGAAGAACTTAAA
CCGACCCCGGAAGGCGATCTGGAAATTCTGCTGCAGAAAGCGCTGCCTGATGATGTGCTGGCAAACGCGTTTCAGATTAGCCGCTTTTATACCG
TGATTAGCAGCCTGAAAGGCACCGGGAATTTAGAACTGGTTGCGGTCCGTACTAACGACAATGCCCAGATTAGCCCGCTGGCAGGCCGTCTAGA
ACCCGGTGACATTCCTGCGTCTCATCACCATCATCATCACCATTAATAACTCGAG

Construct 3-2: Construct with optimized mRNA secondary structure coding for a GST primer, concatenated 19 final peptide biomarkers and poly-H with an expected product 29 kDa. One copy of each one of the 19 final peptide biomarkers are placed in this order: P14-P7-P3-P5-P18-P9-P4-P17-P10-P2-P1-P16-P15-P12-P6-P19-P8-P11-P13

CATATGTCTCCTATTCTAGGTTATCCGGGTGGTAGTCGCGGCAGCGAAGAAGAAGGCGATATTACCAACCCGATTAACCTGCGCGGCGGCCTGG
AACCGATCAACTTTCAGACCGCAGCAGATCAGGCACGCTACCTGGGTTATCTGGAACAGCTGCTGCGCACCCCGGAAGTGGATGATGAAGCGCT
GGAAAAACTGAACGCGCTGGAACCGACCAACCGCTTTTATACCGTGATTAGCAGCCTGAAAATTGATGCACTGAATGAAAACAAAGCGCTGCCG
GATGATGTGCTCGCAAACGCCTTTCAGATTTCTCGCTCGGCAGGCTGGAACATTCCGATTGGCACCCTGATTCATCGCCACCAAGGCCTGCCCC
AGGAAGTGCTGAACGAAAACCTGCTGCGCTTTTTTGTGGCTCCATTTCCAGAAGTATTTGGCAAAGCGGATATTTATACCGAGCAAGTTGGCCG
CGGTACCGGGAACTTGGAGCTGGTTGCAGTGCGCACCGTGATTGTAGAAGCGCCGATTCATGGCCTGAAAGTGTATGTGGAAGAACTGAAACCG
ACCCCGGAAGGCGATCTGGAAATTCTGCTGCAGAAAACCAACGACAATGCGCAGATTTCCCCCTTAGCGGGCCGCCATATTGCGACCAACGCAG
TCCTGTTTTTTGGCCGCAACGTGAACTTTGATGGTGAAATTCTGAAAGGCTCCGAAGAAGAGGATATTACGAACCCGATTAACCTTCGCCTAGA
ACCAGGAGATATCCCGGCGAGCCATCACCATCACCATCACCATTAATAACTCGAG

Construct 3-3: Construct with optimized mRNA secondary structure coding for a GFP primer, concatenated 19 final peptide biomarkers and poly-H with an expected product 29 kDa. One copy of each one of the 19 final peptide biomarkers are placed in this order: P11-P6-P8-P7-P2-P10-P19-P14-P3-P18-P9-P13-P15-P12-P17-P16-P1-P4-P5

CATATGGCTAGCAAAGGAGAAGAACTTTTCACTGGACGTAACGTGAACTTTGATGGGGAGATCTTGAAAGTGTATGTGGAAGAACTGAAACCGA
CCCCGGAAGGCGATCTGGAAATTCTGCTGCAGAAACATATTGCGACCAACGCAGTTCTGTTTTTCGGTCGTGGTGGCCTCGAACCTATTAACTT
TCAGACCGCAGCAGATCAAGCGCGCCATCAGGGTCTGCCGCAGGAAGTGCTGAACGAAAACTTGCTGCGTTCTGCGGGCTGGAACATCCCTATC
GGCACCCTGATTCATCGCACCAACGATAACGCGCAGATTTCACCGCTGGCCGGACGTGGATCGGAAGAAGAAGGCGATATTACCAACCCGATTA
ACCTGCGCTACCTGGGCTATTTAGAACAGCTGTTGCGCCTGAACGCACTGGAACCGACCAACCGCTTTTATACCGTGATTAGCAGCCTGAAAGG
TTCGGAAGAAGAGGATATTACCAACCCGATTAACCTGCGCGGCACCGGTAACCTGGAACTAGTAGCAGTACGCACGGTGATAGTGGAAGCGCCC
ATTCATGGCCTGAAAGCCCTGCCCGACGATGTGCTGGCGAACGCGTTTCAGATTAGCCGCGCGGATATTTATACGGAACAGGTGGGCAGGTTTT
TTGTGGCGCCGTTTCCGGAAGTGTTTGGCAAAATTGACGCTCTGAATGAGAATAAGACCCCGGAAGTGGATGATGAAGCGCTGGAAAAACTAGA
ACCAGGAGATATCCCGGCGAGCCATCACCATCACCATCACCATTAATAACTCGAG

Construct 3-4: Construct with optimized mRNA secondary structure coding for a GFP primer, concatenated 19 final peptide biomarkers and poly-H with an expected product 29 kDa. One copy of each one of the 19 final peptide biomarkers are placed in this order: P11-P17-P1-P14-P7-P9-P19-P16-P10-P13-P8-P15-P18-P12-P4-P6-P3-P5-P2

CATATGGCTAGCAAAGGAGAAGAACTTTTCACTGGACGTAACGTGAACTTTGATGGCGAAATTCTGAAAGCGCTCCCCGATGATGTGCTGGCCA
ACGCGTTTCAGATCAGCCGCTTTTTTTGTAGCACCGTTTCCGGAAGTGTTTGGCAAAGGCAGCGAAGAAGAAGGCGATATTACCAACCCGATTAA
CCTGCGCGGCGGCCTGGAACCGATTAACTTTCAGACCGCAGCAGATCAGGCGCGCTTTTATACCGTGATTAGCTCACTGAAAACCAACGACAAT
GCACAGATCTCTCCCCTCGCTGGTCGCGCAGATATTTATACCGAACAGGTCGGGCGCTCAGCGGGCTGGAATATTCCGATTGGTACCTTAATTC
ATCGCGGCTCTGAAGAGGAAGATATCACCAACCCGATCAATCTGCGTCATATTGCCACCAACGCAGTTCTTTTTTTTGGGAGAGGGACGGGTAA
CCTGGAACTGGTAGCGGTGCGACTCAATGCCCTGGAACCCACGAACCGCACCGTCATTGTGGAAGCGCCGATTCATGGTCTGAAAATCGATGCG
TTGAACGAAAACAAAGTGTACGTGGAAGAACTTAAACCGACCCCGGAAGGCGATCTGGAAATTCTGCTGCAGAAATATCTGGGCTACCTGGAAC
AGCTGCTGCGCACCCCGGAAGTGGATGATGAAGCGCTGGAAAAACATCAGGGCCTGCCGCAAGAAGTGCTGAACGAAAACCTGCTTCGCCTAGA
ACCCGGAGATATCCCGGCGAGCCATCACCATCACCATCACCATTAATAACTCGAG